Three Essays on Household Residential Sorting

by

Zhaohua Zhang

A dissertation submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Auburn, Alabama
August 6, 2016

Keywords: Residential location Sorting, Racial Segregation, Household heterogeneity, Crime, Landfill, Environmental Risk

Approved by

Diane Hite, Chair, Professor of Agricultural Economics
Valentina Hartarska, Professor of Agricultural Economics
Denis Nadolnyak, Associate Professor of Agricultural Economics
Alan Seals, Assistant Professor of Economics

Abstract

This dissertation includes three essays that analyze factors affect residential sorting across neighborhood. Chapter 1 estimates the effect of crime on household location choice using a two-stage residential sorting model which incorporates the effect of mobility cost. The results from the second stage show that people are willing to pay more to move to a location with lower violent crime occurrences, but also willing to pay more to move to a place with higher property crime. When recovering the willingness to pay (WTP) for the two types of crime using elasticities, the results show that people are willing to pay \$651 and \$977 for a one hundred unit decrease in violent crime and \$23 and \$27 for a one hundred unit increase in property crime for 2005 and 2010 respectively. The difference-in-difference results show that people are willing to pay less to move to a location in which the police number increases, and pay more to move to a location where the crime rate decreases while police force increases.

Chapter 2 analyzes whether or not, and to what degree local environmental risk impact household residential location choice. Employing a two-stage horizontal sorting model, the results indicate that black households are willing to pay \$3438 more for a 1% increase in the faction of black than white households, and households of other races would like to pay \$8613 more for a 1% increase in the fraction of same race neighbors than white households. With each \$10000 increase in household income, household's marginal willingness to pay increases by \$591 for a decrease of 1000 pounds of releases in the neighborhood. The counterfactual

simulation of turning off tastes over environmental risk shows that differential preferences for environmental risk by race serve to segregate households.

Chapter 3 analyzes how environmental disamenities affect residential location choices using an equilibrium sorting model. The empirical analysis indicates that households are heterogeneous in their willingness to pay for housing and neighborhood characteristics, which shaped the way that households sort across neighborhood. Based on results from the first stage estimation, poor households are more likely to select houses closer to landfills, and black households are more probably to choose households located near a demolish landfill when keeping distance fixed. Counterfactual simulation results indicate that when switching off heterogeneous preferences for landfill disamenities there is little impact on housing consumption of white, but for black and poor households, I see more notable changes.

Acknowledgments

I would like to express the deepest appreciation to my academic advisor, Dr. Diane Hite, for his continuous support, academic advice, understanding, patience and encouragement during my Ph.D. study.

Besides my advisor, my sincere thanks also go to Dr. Valentina Hartarska, Dr. Denis Nadolnyak, and Dr. Alan Seals for serving as my advisory committee members and giving me insightful comments and suggestions. I also would like to thank Dr. Duha T. Altindag to serving as the outside reader for my dissertation. I would also like to thank all my professors and colleagues in the Department of Agricultural Economics and Rural Sociology for their instructions and help during my graduate study.

Last but not least, I also wish to extend my thanks to my family: my parents, my husband and my son for supporting me spiritually throughout writing this dissertation and my life in general.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ACS	American Community Survey
EPA	Environmental Protection Agency
FBI	Federal Bureau of Investigation
IPUMS	Integrated Public Use Microdata Series
MSA	Metropolitan Statistical Area
MWTP	Marginal Willingness to Pay
WTP	Willingness to Pay

Chapter 1 . House Value, Crime and Residential Location Choice

1. Introduction

The study of residential location choice has captured the interest of researchers in a diverse range of disciplines, including economists, geographers and sociologists. As a result, different methods have been introduced to do the analysis. After Rosen's (1974) seminal paper, the hedonic price model became one of the most popular methods to analyze housing market issues. Charles Tiebout's (1956) introduced the Tiebout model and after that the Tiebout model was widely used in analyzing the provision of public goods. The method used in this study, which is developed by Bayer (2009), is an extension of the traditional sorting model that introduces mobility cost.

People select the place where they live for different reasons, such as a new job opportunity or to be together with their family. However, when making decisions, public goods and amenities (e.g. clean air, school quality, and lower crime rates) are also characteristics that people care about. As a result, the dwelling choice of households involves trade-offs among a series of factors that affect the utility of households. For example, when people move from a place with a higher crime rate to a place with a lower crime rate, it is true that it is a much safer place for people to live but they may also experience a decrease in their wage and an increase in housing prices. To maximize their utility, people must compare these trade-offs carefully. Changes in housing prices and income reflect the willingness to pay for local amenities. Though I can obtain the implicit price of the local amenities by using the hedonic price model, the assumption of this model overlooks an important problem: the cost of migration and the effect of income change. Moving to a new place not only costs money but also includes psychic cost due to leaving behind our cultural roots. Thus, if we do not take these costs into account, the value of an amenity will be overstated. While previous sorting models that analyze the impact of public goods on household location choice ignore the cost of migration, in this study I follow Bayer et al. (2009) by including moving costs and modeling choices across MSAs to implement the analysis. This study concentrates on the effect of crime occurrences on household residential location choice. As in previous papers, this study analyzes the relationship between dwelling location choice and public goods using observed behavior in the housing markets. What is different is that this study estimates both short and long-run migration impacts and adds a police force variable to control for crime impacts using difference in difference methodology.

In this study, I model the dwelling decision as a choice among different metropolitan statistical areas (MSAs) taking potential income, house prices, moving cost, the number of crime occurrences and other location-specific characteristics into account. The discrete choice model is used to infer different utilities for household that live in different MSAs in 2005 and 2010. Then I regress these utilities on the number of crime occurrences and local characteristics that varying by MSA, in order to find the marginal willingness to pay (MWTP) for moving to a location with lower crime and compare the results for 2005 and 2010. After that, I use a difference in difference model to estimate the impact of changes of police and crime rates on individual's MWTP.

The remainder of the study proceeds as follows. The next section provides the literature review. The details of the analytical framework and methodology are shown in section 3. Section

4 describes the data and variables I use in the models and is followed by the analysis of estimation results. The final part is the conclusions.

2. Literature Review

There is a vast body of literatures analyzing the relationship between house prices, crime and household residential location choice. In this section I discuss some of the previous literature addressing some of the same topics as in this study.

Household location choices have been of continuing interest to economists for decades. Rosen (1974) first introduced the basic theory for analyzing housing market prices using the hedonic price model. However, the hedonic price model has methodological issues such as identification and endogeneity problems. Chay and Greenstone (2005) showed problems with the identification and consistent estimation of the hedonic price model. They were interested in endogeneity of the pollution variable and introduced an instrumental variable approach to estimate it consistently. They also showed that if there existed heterogeneity in preference functions, endogeneity existed when sorting by house purchasers with different pollution levels. Anselin and Gracia (2008) discussed spatial autocorrelation and heteroscedasticity in the error terms when estimating hedonic models of house prices. This study faced the problem of mismatch between the spatial support of the explanatory variable, a pollution measure collected at a finite set of monitoring stations, and the dependent variable, the price observed at the location of the house sales transaction. To deal with this problem, the authors used a spatial econometric approach and included a spatially lagged dependent variable in the hedonic specification.

Since the hedonic price model has its shortcomings, a new method, the "discrete choice model", was introduced to study the residential location choice problem. The earliest attempt to apply discrete choice theory to residential location analysis was McFadden (1978). He provided a solution to the problem of modeling disaggregate choice of housing location when the number of disaggregate alternatives was impractically large and when the presence of a structure of similarities between alternatives invalidates the commonly used jointly multinomial Logit choice model. Through the choice process of individuals, the population is sorted into optimum communities according to the tastes of residents. Bayer et al. (2002) presented a new equilibrium framework for analyzing economic and policy questions related to the sorting of households within a large metropolitan area which incorporates choice-specific unobservables to identify household preferences over choice characteristics. Bayer et al. (2007) developed a framework for estimating household preferences for school and neighborhood attributes in the presence of sorting, using restricted access Census data from a large metropolitan area. This study introduced a boundary discontinuity design to a heterogeneous residential choice model, addressing the endogeneity of school and neighborhood characteristics. Yu et al. (2012) analyzed the relationship between residential location choice and household energy consumption behavior, using a joint mixed Multinomial Logit-Multiple Discrete-Continuous Extreme Value model by controlling for self-selection. Recent research by Duijn and Rouwendal (2013) investigated the impact of cultural heritage on the attractiveness of cities by analyzing the location choice of households applying a residential sorting model. In this study, they also used spatial econometric techniques to extend the residential sorting model to incorporate the effect of amenities of the nearby locations. Kuminoff et al. (2013) built unemployment into a model of sorting across the housing and labor markets to evaluate the welfare effect of a prospective regulation that would improve environmental quality while simultaneously generating layoffs. The sorting model is used to analyze the effect of different factors on house prices and residential location choice. The method used in this study follows Bayer’s (2009) model which extends McFadden’s (1978) model by introducing mobility cost into the model.

Many researchers suggest that dwelling location decisions and house prices can be affected by crime rates. Interestingly, different studies come to different conclusions. Cullen et al. (1999) found negative relationships between them. The study of Lynch and Rasmussen (2001) estimates the impact of crime on house prices using data on over 2800 house sales in Jacksonville, FL. But the results showed that cost of crime had virtually no impact on house prices overall, but that homes were highly discounted in high crime areas. Gibbons and Machin (2008) considered the role of local amenities and disamenities in generating the variation of house prices within urban areas, focusing on three highly policy-relevant urban issues--transport accessibility, school quality, and crime. A recent study by Ihlanfeldt and Mayock (2009) studied the effect of neighborhood crime on housing values, finding that a 10 percent increase in violent crime within a neighborhood was found to reduce housing values by as much as 6 percent. Ihlanfeldt and Mayock (2010) utilized a nine-year panel of crime for Miami-Dade County at the neighborhood level to analyze the impact of crime on house prices. They found that house buyers were willing to pay nontrivial premiums for housing located in neighborhoods with less aggravated assault and robbery crime, with elasticities of house value with respect to aggravated assault crime and robbery crime of -0.152 and -0.111, respectively. Frischtak and Mandel (2012) used a recent policy experiment in Rio de Janeiro, the installation of permanent police stations in low-income communities, to quantify the relationship between a reduction in crime and the change in the prices of nearby residential real estate. Although these papers analyzed the impact of crime on house prices, I did not find any paper that use sorting model to analyze the effect of crime on household dwelling location choice. Thus, I apply a sorting model to analyze the effect of crime on residential location choice in my study and also compare the results with the conventional hedonic price model which has been commonly used in the literatures.

3. Methodology

3.1 Conceptual Model

In theoretical models of residential sorting, the residential location choice of households is closely related to the demand for local public services, such as lower crime rate. In this subsection, I start with individual choice behavior, and then introduce the model of residential sorting. Individual choice behavior is modeled by postulating a utility function U whose value is determined by the consumption of a composite commodity C and the characteristics of house H. The quantity of amenity ("the reduced number of crime occurrence") in location j is defined as X_j and the moving cost of settling in location j is M_j. Since moving from one place to another does not only cost money but also produces psychic cost of leaving behind one's cultural roots, moving cost is introduced into the utility equation instead of being introduced into the budget constrain function. To keep the theoretical model simple and capture the fixed moving cost in location j, following Bayer (2009), this study assumes that all people are born in the same place. When making decisions, individuals choose their living location simultaneously and each individual chooses her location j to maximize her utility subject to a budget constraint:

$$(1) \qquad \max_{\{C,H,X_j\}} U(C,H;X_j,M_j) \quad s.t.\ C+\rho_j H = I_j$$

where I_j is income in location j; ρ_j is the price of housing in location j; and the composite good is available in continuous quantities at a unit price normalized to 1. Individuals maximize their utility by determining the values of C and H. After substitution of the optimal value of these variables into the utility function (1), we can get the indirect utility function V:

$$(2) \qquad V(I_j,\rho_j;\ X_j,M_j) \equiv \overline{V}$$

$\overline{V}$ denotes indirect utility. Following Roy's identity, the Marshallian demand function for house can be expressed as $H = -V_\rho/V_I$, where V_ρ and V_I are partial derivatives of indirect utility function with respect to house prices and income respectively. Taking the total derivative of equation (2) and substituting for $H = -V_\rho/V_I$ we can get the implicit price of the amenity as follows:

(3) $$P^* = H\frac{d\rho}{dX} - \frac{dI}{dX} - \frac{V_M}{V_I}\frac{dM}{dX}$$

Where V_M represents the partial derivative of the indirect utility function with respect to mobility cost. Thus, P^* is the MWTP. From equation (3) we know that if mobility is costless ($V_M = 0$) or mobility cost is constant under which condition $\frac{dM}{dX} = 0$, this equation is the same as the traditional hedonic price model. If mobility is costless or mobility cost is constant, or we know what M really is, we can get the MWTP for amenities. However, in reality, we could not observe M and if we want to get the MWTP for the amenity it is necessary to consider moving cost, and a different method needs to be introduced. Following Bayer (2009), I start with the following utility function assuming that individual i lives in location j, and consumes quantities C_i goods and housing type $H_i$respectively:

(4) $$U_{i,j} = C_i^{\beta_C} H_i^{\beta_H} X_j^{\beta_X} e^{M_{i,j}+\xi_j+\eta_{i,j}}$$

where, X_j denotes the amenity in location j; $M_{i,j}$ measures the long-run and short-run (dis)utility of migration associated with moving from person i's birth location to destination j; ξ_j contains unobserved characteristics of location j and $\eta_{i,j}$ represents an individual-specific diosyncratic component of utility which is assumed to be independent of mobility costs and location characteristics. β_C, β_H, and β_X are parameters associated with the consumption of goods, house and the local amenity. Applying the budget constraint in equation (1), differentiating with respect to H_i, and rearranging we can get the housing expenditure as follows:

(5) $$H_{i,j}^* = \frac{\beta_H}{\beta_H + \beta_C} * \frac{I_{i,j}}{\rho_j}$$

"*" here represents the optimal result from utility maximization function. Since people do not explicitly pay for consumption of the local amenity, $\beta_H/(\beta_H + \beta_C)$ represents the share of income spent on housing. Substituting equation (5) into (4) and using the budget constrain, the indirect utility function can be expressed as follows:

(6) $$V_{i,j} = I_{i,j}^{\beta_I} e^{M_{i,j} - \beta_H ln\rho_{j+}\beta_X lnX_{j+}\xi_j + \eta_{i,j}}$$

where $\beta_I = \beta_H + \beta_C$, which is the parameter associated with income. MWTP for the amenity equals the marginal rate of substitution between X_j and income and for person i, MWTP can be expressed as $MWTP_i = \frac{\beta_X}{\beta_I}\frac{I_{i,j}}{X_j}$. However, in reality, we just know the income people get from the location where they live and work, and as a result we need to estimate income that people would get from alternative locations. Here I express it as: $I_{i,j} = \hat{I}_{i,j} + \varepsilon_{i,j}^I$, where $\hat{I}_{i,j}$ is the predicted income for person i in all the j locations and $\varepsilon_{i,j}^I$ is the error term. Substitute this into function (6) and taking logs we can get the following equation:

(7) $$lnV_{i,j} = \beta_I ln\hat{I}_{i,j} + M_{i,j} + \theta_j + \nu_{i,j}$$

where

(8) $$\theta_j = -\beta_H ln\rho_j + \beta_X lnX_j + \ \xi_j$$

and

(9) $$\nu_{i,j} = \beta_I \varepsilon_{i,j}^I + \eta_{i,j}$$

where θ_j is a location-specific term. I assume that all the random terms are independently identically distributed with extreme value type I distribution (McFadden, 1973). For convenience, I divide all the variables by β_I denoted as tildes, for example $\tilde{\theta}_j = \frac{\theta_j}{\beta_I}$ and then the choice probabilities of households to maximize their utility can be shown as:

$$(10) \qquad p\left(ln\tilde{V}_{i,j} \geq ln\tilde{V}_{i,k}, \forall k \neq j\right) = \frac{e^{\sigma(ln\hat{I}_{i,j}+\tilde{M}_{i,j}+\tilde{\theta}_j)}}{\sum_{n=1}^{j} e^{\sigma(ln\hat{I}_{i,n}+\tilde{M}_{i,n}+\tilde{\theta}_n)}}$$

where $\sigma = 1/\beta_I$ is a scaling parameter. This is the multinomial Logit model which is estimated using maximum likelihood. Household select location j as long as $ln\tilde{V}_{i,j} \geq ln\tilde{V}_{i,k}$ and p represents the probability household i chooses location j. $\tilde{\theta}_j$ is regarded as parameters in this model. The focus of estimating equation (10) is to get $\tilde{\theta}_j$, which will be used in the second stage. In the second stage, the estimated $\tilde{\theta}_j$ is regressed on crime rate and other location characteristics. From equation (8), the equation for the second stage can be written as follows:

$$(11) \qquad \tilde{\theta}_j = -\tilde{\beta}_H ln\rho_j + \tilde{\beta}_X lnX_j + \tilde{\xi}_j$$

$\tilde{\beta}_H = \frac{\beta_H}{\beta_I}$ represents the share of income spent on house and $\tilde{\beta}_X = \frac{\beta_X}{\beta_I}$ represents the share of income spent on other goods. As shown in previous part, for person i, $MWTP_i = \frac{\beta_X}{\beta_I}\frac{I_{i,j}}{X_j}$. Thus, $\tilde{\beta}_X = MWTP_i \times \frac{X_j}{I_{i,j}}$ from which the WTP for lower crime rate can be estimated.

3.2 Econometric implementation

The underlying assumption of the second-stage regression is that house prices are uncorrelated with unobserved characteristics of residential locations. However, the observed prices are often correlated with the unobservable attributes. For example, house prices may be affected by the prices of the nearby houses and if we ignore the endogeneity of house prices, the estimation results will be biased. Thus, to eliminate the correlation between house prices and unobserved location characteristics and the correlation between amenity and unobserved local attributes, this study followed Chay and Greenstone estimating equation (11) by moving $-\tilde{\beta}_H ln\rho_j$ to the left and then equation (11) can be written as:

(12) $\tilde{\theta}_j + \tilde{\beta}_H ln\rho_j = \tilde{\beta}_X lnX_j + \tilde{\xi}_j$

However, to implement the residential sorting model, the following problems still need to be solved. The first problem is about how to get the "price of housing service". The following functions is used to estimate house prices taking the characteristics of individual house into account:

(13) $lnP_{i,j,t} = ln\rho_{j,t} + \omega_t h_{it} + \varepsilon_{i,j,t}$

where $P_{i,j,t}$ is the value of the house i in location j; h_{it} represents the characteristics of house i in time t; $\rho_{j,t}$ is a scaling parameter; ω_t is the parameter that needs to be estimated and $\varepsilon_{i,j,t}$ is the error term. Since the index of "housing services" could be defined as $H_{it} = e^{\omega_t h_{it}}$ which can be estimated using parameter ω_t, we can use $\rho_{j,t}$ as the measurement of the effective "price of housing service" which provides a consistent measure of the true prices of house. The house characteristics in this analysis contain the number of rooms, the number of bedrooms, the number of housing units, the age of the house, and the acres of the house.

The second problem relates to income. Since we can not observe the income of individuals in every location, the following equation is used to estimate the MSA level income for each individual (Bayer 2009):

(14) $$lnI_{i,j,t} = \alpha_{0,j,t} + \alpha_{white,j,t} white_{i,t} + \alpha_{male,j,t} male_{it} + \alpha_{age>60,j,t} age > 60_{i,t} + \alpha_{hsdrop,j,t} hsdrop_{i,t} + \alpha_{somecoll,j,t} somecoll_{it} + \alpha_{collgrad,j,t} collgrad_{i,t} + \varepsilon^I_{i,j,t}$$

Where $white_{i,t}$ is a dummy variable indicating whether the race of person i is white; $male_{it}$ is a dummy variable which equals 1 if the gender of person i is male; $age > 60_{i,t}$ is a indicator which equals 1 if a person is older than 60; $hsdrop_{i,t}$ is a dummy variable that equals 1 if person i drop out in high school; $somecoll_{it}$ denotes person i get some college education that less than four years and $collgrad_{i,t}$ equals 1 if person i get college degree and higher. $\varepsilon^{I}_{i,j,t}$ is the error term and the α are coefficients we need to be estimated. By estimating the above function, we can generate the predicted value of income for each individual in all locations.

Then, it comes to mobility cost. The following function is used to measure the mobility (Bayer et al. 2009):

(15) $$\widetilde{M}_{i,j,t} = \tilde{\mu}_s d^s_{i,j,t} + \tilde{\mu}_r d^r_{i,j,t} + \tilde{\mu}_m d^m_{i,j,t}$$

Where $\widetilde{M}_{i,j,t}$ represents the moving cost which is a function a series of dummy variables. $d^s_{i,j,t}$ is a dummy variable which equals 1 if location j is not in the state where individual i was born (=0 otherwise); $d^r_{i,j,t}$ is an indicator whether a person lives in the same region as his/her birth region[1]; and $d^m_{i,j,t}$ equals 1 if location j is not in the macro-region[2] where individual i was born. In this equation, $\tilde{\mu}_s$, $\tilde{\mu}_r$ and $\tilde{\mu}_m$are all parameters. Following Bayer, this equation represents long-run utility cost and captures the psychic cost due to leaving behind one's cultural roots. In order to compare this long-run utility cost, I also estimate the mobility cost function which captures the short-run utility. The short-run mobility cost function is defined as follows:

(16) $$\widetilde{M}_{i,j,t} = \tilde{\alpha}_s m^s_{i,j,t} + \tilde{\alpha}_r m^r_{i,j,t} + \tilde{\alpha}_m m^m_{i,j,t}$$

---

[1] Regional Definitions: (1) New England (CT, ME, MA, NH, RI, VT), (2)Middle Atlantic (NJ, NY, PA), (3) East North Central (IL, IN, MI, OH, WI), (4) West North Central (IA, KS, MN, MO,NE, SD, ND), (5) South Atlantic (DE, DC, FL, GA, MD,NC, SC, VA, WV), (6) East South Central (AL, KY, MS, TN), (7) West South Central (AR, LA, OK, TX), (8) Mountain (AZ, CO, ID, MT, NV, NM, UT,WY), and (9) Pacific (AK,CA, HI, OR, WA).

[2] There are four macro-regions defined by US census bureau: (1) Northeast (New England, Middle Atlantic), (2) Midwest() East North Central, West North Central), (3) South(South Atlantic, East South Central, West South Central), (4) West(Mountain, Pacific).

In equation (16) $\widetilde{M}_{i,j,t}$ is still the mobility cost. $m^s_{i,j,t}$ is a dummy variable which equals 1 if the current state that people is not the one they live one year before; $m^r_{i,j,t}$ indicates whether people live in the same region as they lived one year before, and $m^m_{i,j,t}$ is also a dummy variable which equals 1 if location j is not in the same macro region as people lived one year ago. All the $\tilde{\alpha}$ are parameters that need to be estimated. The difference between equation (15) and equation (16) is that equation (16) represents the real moving cost in the short-run while equation (15) captures both the long-run moving cost and the psychic cost of leaving behind the cultural root.

By investigating previous studies I find that researchers used either total crime or a single type of crime to measure crime. In this study, by adding up the number of crimes that occurred, I estimate the effect of property and violent crime on people's WTP separately. Violent crime here includes murder and non-negligent manslaughter, forcible rape, robbery and aggravated assault. Property crime sums up burglary, larceny-theft and motor vehicle theft.

For the first step, I estimate the parameters of $\tilde{\mu}_s, \tilde{\mu}_r, \tilde{\mu}_m, \sigma$ and $\tilde{\theta}_j$ from the following likelihood function in which we also assume that all the random terms are independently and identically distributed with extreme value type I distribution[3]:

$$(17) \qquad L(\tilde{\mu}_s, \tilde{\mu}_r, \tilde{\mu}_m, \sigma, \tilde{\theta}_t) = \prod_t \prod_i \prod_{j=1}^{J} \left[\frac{e^{\sigma(ln\hat{I}_{i,j,t} + \tilde{\mu}_s d^s_{i,j,t} + \tilde{\mu}_r d^r_{i,j,t} + \tilde{\mu}_m d^m_{i,j} + \tilde{\theta}_{j,t})}}{\sum_{n=1}^{J} e^{\sigma(ln\hat{I}_{i,n,t} + \tilde{\mu}_s d^s_{i,j,t} + \tilde{\mu}_r d^r_{i,j,t} + \tilde{\mu}_m d^m_{i,j} + \tilde{\theta}_{n,t})}} \right]^{x_{i,j,t}}$$

where $x_{i,j,t}$ is an indicator that equals one if household i observed in year t chooses to live at location j. All the other symbols are the same as equations (10) and (15). Let $C_{j,t}$ denote the crime rate in location j period t and $R_{j,t}$ represents other location characteristics, then

[3] Here, I just show the likelihood function for long-run mobility cost for the sake of compact. The likelihood function for short-run mobility cost can be get by replacing $\tilde{\mu}_s, \tilde{\mu}_r, \tilde{\mu}_m$ with $\tilde{\alpha}_s, \tilde{\alpha}_r, \tilde{\alpha}_m$ and replacing $d^s_{i,j,t}, d^r_{i,j,t}, d^m_{i,j}$ with $m^s_{i,j,t}, m^r_{i,j,t}, m^m_{i,j,t}$.

equation (12) can be rewritten as the following equation, which will be estimated in the second step:

(18) $\tilde{\theta}_j + \tilde{\beta}_H ln\rho_j = \tilde{\beta}_C lnC_j + \tilde{\beta}_R lnR_j + \tilde{\xi}_j$

In equation (18), the value of $\tilde{\theta}_j$ is obtained from the first stage by using maximum likelihood estimation, and $ln\rho_j$ is the "house service price" was obtained from the estimation of equation (13). C_j represents crime in location j and R_j represents other location characteristics. $\tilde{\beta}_C$ and $\tilde{\beta}_R$ are coefficients to estimate. $\tilde{\beta}_H$ in equation (18) is the share of income spent on housing. To estimate the share, I use the annual average 30-year fixed mortgage rate of 2005 and 2010 which is 6.01% and 5.08% respectively[4]. Using this rate I can estimate the annual value for each house using the following equation:

(19) $$AV = \frac{TV * R}{1 - (1 + R)^{-n}}$$

Where AV represent the annual value of the house; TV is the total value of the house; R represents the 30-year fixed mortgage rate and n is the total period of installment which is 30 years in this case. When the annual value for each house is calculated, the share of income spent on housing can be expressed by the ratio of annual house value to income. The median share of income spent on housing is used in my study. [5]

3.3 Difference in Difference Analysis

The innovation of this study from other papers is that I introduce a difference in difference analysis to analyze the effect of the change in crime rate and police numbers on individual's WTP. The estimation results of sorting model for 2005 and 2010 are obtained from the previous analysis. However, comparing the two results, we could not decide whether the

[4] These values can be get from HSH.com.
[5] The median share of income for the whole micro data sample is 0.36 and 0.29 for the year 2005 and 2010 respectively. Thus, $\tilde{\beta}_H$ in equation (18) equals to 0.36 for 2005and equals 0.29 for 2010.

change of WTP from 2005 to 2010 was a result of a change in crime incidence during this period, or if other factors influenced MWTP. The obvious evidence is that the occurrence of crimes is related closely to the number of police in a location. Thus, by combining the data from 2005 and 2010, I use both the change in crime occurrence and the change in police force size from 2005 and 2010 as treatment to do a difference in difference analysis.

In reality, an increase in police number does not indicate a decrease in the crime rate, thus in this study I use two treatments. One is the increase in police numbers of per thousand people and the other is the decrease in crime rate from 2005 to 2010. The crime rate is measured by the ratio of the total number of crime (including both property crime and violent crime) to population in each location. Difference in difference is measured with a minor change in the second stage estimation as follows:

$$(20) \qquad \tilde{\theta}_j + \tilde{\beta}_H ln\rho_j = \delta_1 T + \delta_2 D_{1j} + \delta_3 D_{2j} + \delta_4 D_{1j} * T + \delta_5 D_{2j} * T + \delta_6 D_{1j} * D_{2j}$$
$$+\delta_7 D_{1j} * D_{2j} * T + \tilde{\beta}_R lnR_j + \tilde{\xi}_j$$

where $\tilde{\theta}_j$, $\tilde{\beta}_H$, $\tilde{\beta}_R$, lnR_j and $ln\rho_j$ are defined as previously. T is a dummy variable which equals to 1 for 2010; D_{1j} equals 1 if the police number in location j increased from 2005 to 2010 and D_{2j} indicates whether the crime rate in location j decreases from 2005 to 2010. In equation (20), the "δ"s are the parameters to be estimated, with δ_1 capturing the time trend; δ_2, δ_3 and δ_6 capture treatment group specific effects and δ_4, δ_5 and δ_7 represents the true treatment effect which is the interest of the difference in difference analysis. To avoid endogeneity, I still move house prices to the left hand side.

In equation (20), the left hand side variable is from the sorting model, in order to make comparison, I also estimate the difference in difference effect using a traditional hedonic price model in which the left hand side variable is the house value. The hedonic price model is defined as follows:

$$(21) \quad lnP_{i,j,t} = ln\rho_{j,t} + \omega_t h_{it} + \delta_1 T + \delta_2 D_{1j} + \delta_3 D_{2j} + \delta_4 D_{1j} * T + \delta_5 D_{2j} * T + \delta_6 D_{1j} * D_{2j}$$
$$+\delta_7 D_{1j} * D_{2j} * T + \varepsilon_{i,j,t}$$

Where $lnP_{i,j,t}$ is the logarithmic form of the house value in location j of time period t; h_{it} represents the characteristics of the house i in time t; $\rho_{j,t}$ is a scaling parameter and $\varepsilon_{i,j,t}$ is the error term. Other variables are defined the same as those in equation (20).

4. Data sources

Date used in this study comes from several sources, all of which are publicly available. The choice set used to analyze individuals' residential decisions in this study is the metropolitan statistical areas (MSAs) in the United States. The individual income prediction and the house prices estimation are also at MSA level. A map of all the MSAs in the United States is shown in figure 1.1. Though figure 1.1 shows that there are many MSAs that are not contiguous to each other, most MSAs share the same border. Data used to estimate discrete choice model of residential location choice, individual income and housing service price are obtained from the American Community Survey (ACS) 2005 and 2010 sample.

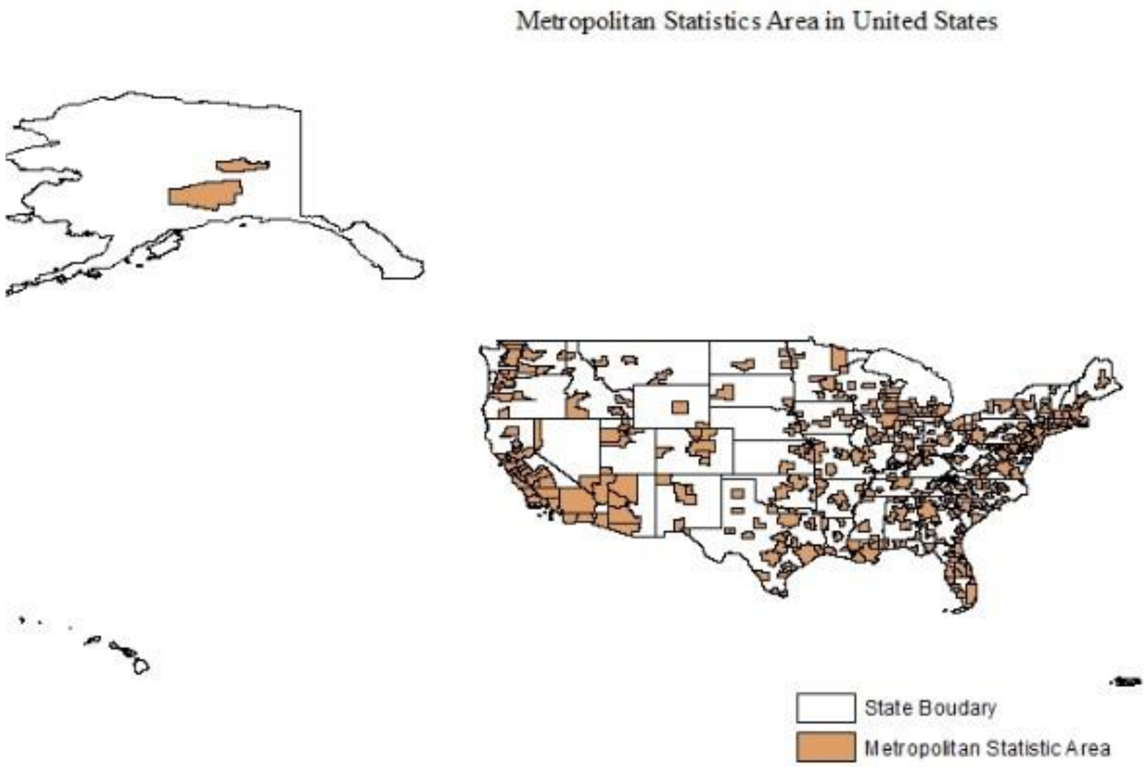

Figure 1.1 Metropolitan Statistics Areas in United States

The ACS sample provides a variety of data at MSA level including individual information as well as dwelling characteristics. In the estimation of location specific incomes, I consider the household head as the decision maker, and during my analysis, all householder attributes are relevant to household head. Variables used to estimate housing service price include house value, the number of rooms, the number of bedrooms, the age of the house, the units of the structure and the acres of the house. Variables used to predict income include the sex, age, race, and education attainment of the household head. All these data can be downloaded from Integrated Public Use Microdata Series (IPUMS) which is a project dedicated to collecting and distributing United States census data. The data used to estimate mobility cost are calculated from the data describing the birth state of household head and the location where they live now, which can be also obtained from IPUMS. The crime data used in my study are obtained from the Federal Bureau of Investigation's (FBI) annual report entitled "Crime in the United States". These annual reports include the total reported numbers of violent and property crime incidents. Police force data are also obtained from the FBI’s website. FBI provides the employee data for all the metropolitan counties in each state and the MSA-level employee data can be obtained by combining the county-level data. For the second stage analysis, information about local employment, per capita personal income and population is needed. All these data are obtained from the Bureau of Economic Analysis. After aggregating the data and dropping the MSAs that are not included in any of the above datasets, there are 221 metropolitan statistical areas left. Since the ACS sample is very large, I random select 20,000 observations from the sample to do the analysis.

The list of all the variables used in my study and the summary statistics of all the variables are given in Table A1. The summary statistics show that most of the houses in the sample are smaller than 10 acres and a house with 6 rooms and 3 bedrooms is the most common type. For individual variables, it shows that most household head in the sample are white female and aged older than 60.

5. Estimation Results

5.1 Estimation Results of Incomes and House Prices

By using a two-step strategy, the effect of crime on household dwelling decision, taking mobility cost into account, can be estimated easily. Before carrying out the first step-discrete choice model, I must first estimate individual incomes and house service prices first. In the sample data, we can only observe individual income in the location where he/she lives, thus we need to predict the individual incomes in every location. Using MSA-level population data in 2005 and 2010, the estimation results of location specific mean income described in equation (14) for each year are shown in table A2. For educational attainment variables, the base group includes the individuals with high school degree. For both years, the results show that males earn more money than females after controlling for other variables, and white people earn more than people of other races as expected. People over 60 years old earn less than people who are younger than 60, which is consistent with the reality that many people are retired after age 60. Individuals who dropped out of high school earn less than those with higher education attainment.

Table A3 shows the estimation results of housing service prices described in equation (13). Because in the dataset, the number of rooms and bedrooms is top coded, I create a dummy variable for each number of rooms and bedrooms and the description of the variables is shown in table A1. For the year 2010, the more than 7 rooms do not affect house prices significantly and the effect of the housing unit of boat, tent and van is not statistically significant for either years. The other variables have significant effects on house value in both years. The estimation results show that newer and larger houses yield more housing services. Comparing the housing service price (in logs) in 2005 and 2010, it rose about 36.5% from 2005 to 2010. According to the results, not all the room variables are signed correctly in 2010. I attribute this to the fact that I actually have many correlated measures of size and counts of rooms of different types (total rooms and bedrooms).

5.2 Estimation Results of Residential Sorting Model

When using a sorting model to analyze location choice, there are usually two stages. The first stage is a multinomial Logit model for the personal choice and the second stage is an ordinary least square estimation at the location level. In the first stage, specification of the choice is very important for analyzing the Logit model. The estimation results of the discrete choice equation (17) are shown in table 1.1. For the long-run mobility cost, all results are statistically significant at the 1% level. As expected, living out of individuals' birth state and birth region has negative effects on utility in both 2005 and 2010. Cost continues to increase with living out of one's birth region and macro region. Leaving one's birth state and birth region has almost the same effect on residential location choice in 2005 and 2010. However the cost associated with living out of one's birth macro region increased in 2010. The results for short-run mobility cost show that living in a different state from the one lived in a year before has a significant utility cost in both 2005 and 2010, and the cost continues to increase with living in a different region than one year before. What is different from the long-run mobility cost is that the cost associated with living in a different region as one year before in 2010 is larger than that in 2005. The cost of living in a different macro region as one year before is not statistically significant in 2010. This may be explained by the fact that in the short-run there are less people who change the macro region where they live because of the poor economy.

Table 1.1 Results of Multinomial Logit Estimation

	variable	2005	2010
Long-run	Living out of birth state $\tilde{\mu}_s$	-3.735***	-3.763***
Mobility	Living out of birth region $\tilde{\mu}_r$	-1.347***	-1.305***
Cost	Living out of birth macro region $\tilde{\mu}_m$	-0.501***	-0.540***
	Scaling parameter σ	-0.062	0.009
Short-run	Living in different state as one year before $m^s_{i,j,t}$	-4.499***	-4.487***

Mobility Cost	Living in different region as one year before $m_{i,j,t}^{r}$	-3.789***	-4.156***
	Living in different macro region as one year before $m_{i,j,t}^{m}$	-0.855***	-0.224
	Scaling parameter σ	-0.032	0.043

Note: ***represents statistically significant at %1 level. All data used in the estimation are obtained from 2005 and 2010 American Community Survey data which could be downloaded from the IPUMS website.

What needs to be mentioned here is that the most important thing in the first stage estimation is to get the location fixed effect $\tilde{\theta}_j$ which is not shown in table 1.1[6]. However, when the choice set is large like in this study, the estimation of the Logit model to get $\tilde{\theta}_j$ will become difficult. One of the most commonly used methods is random selection, during which some alternatives are randomly selected from the remaining nonchosen alternatives that the decision maker faces. Using the random selection method, I can estimate parameters for all the observations in the sample. However, as shown in equation (18), the focus of the first stage estimation is to get the fixed effect parameter $\tilde{\theta}_j$ for each location and the larger $\tilde{\theta}_j$ is, the more attractive the location is. In order to get the whole vector of $\tilde{\theta}_j$, Berry et al. (1994) introduced a method to relate market shares to a scalar unobserved choice characteristic. In this study, I apply Berry's method to estimate $\tilde{\theta}_j$ indirectly. Even though the location fixed effect $\tilde{\theta}_j$ represents the preference of people to live in this location, we cannot say that people prefer to live in the location with higher $\tilde{\theta}_j$ because the size of county also affects $\tilde{\theta}_j$. For large MSAs the share of observations who live in these MSAs may be larger than the share of observations who live in the MSAs with a small population. Without controlling for population, according to the rank of $\tilde{\theta}_j$, the most attractive metropolitan area for people to live in both 2005 and 2010 is New York-

[6] There are 221 fixed effects, which are not reported in table 1.1 for the sake of space.

Northeastern, NJ. The least attractive metropolitan area is Iowa City, IA and Alexandria, LA for 2005 and 2010 respectively. However, we cannot make any conclusions without controlling for population. Controlling for population gives us more precise conclusions and the results show that even after controlling for population, New York-Northeastern is still the most attractive location for both year, but Kokomo, IN became the least attractive location for both 2005 and 2010, which is different from the results without controlling the population. Thus, in the analysis, population needs to be taken into account to control for city-size effects.

These MSA-level fixed effects are used in the second stage regression as described in equation (18), and the estimated results for both years are shown in table 1.2. To control population, in the regression, I divide the MSA-level fixed effect by the population in each location. In the estimation, I use the number of property crime and violent crime as the regressors and the focus of the estimation is the coefficient for both regressors. Table 1.2 shows that the effects of both kind of crime are statistically significant and these coefficients represent the elasticities of WTP with respect to property crimes and violent crimes. For both 2005 and 2010, people would like to pay more money to move to a location with lower violent crime occurrence which is consistence with reality. However, for property crime, it is opposite and people are willing to pay more to move to a place with higher property crime occurrence. The same result is also found by previous researchers. Lynch et al. (2001) and Case et al. (2005) showed that the number of violent crimes significantly reduced house values, whereas the number of property crimes had a positive and significant impact on the sales price. This can be explained as higher house prices are more enticing to property crime because the value of the goods inside is expected to be more than that of a lower priced home. Also, locations with higher house prices are richer and more attractive for people to live because people living in richer locations usually have higher income. As a result, people with large amounts of money can spend more on security devices for their home so that non-violent crime does not deter them.

Table 1.2 Results of Second Stage Estimation

Variable	Description	2005		2010	
		value	t-statistic	value	t-statistic
Constant	Intercept	-7.73	-0.99	-3.48	-0.53
Log(ProCri)	Number of property crime	0.03	4.22	0.04	4.93
Log(VioCri)	Number of violent crime	-0.11	-2.45	-0.15	-2.88
Log(Employment)	Fraction of population employed	-3.71	-2.92	-2.58	-2.17
Ln(PerInc)	Per capital personal income	6.52	6.2	5.42	5.06

Note: The estimation results are controlled for the population effect by using the MSA-level fixed effect. Crime data are obtained from FIB 2005 and 2010 crime report.

Since households make decision depending on the trade-off between income, house prices and crime, people are still willing to pay to move to a place with higher house value even though it costs people more to move to a location with high property crime occurrence. It should to be mentioned that the study area in this study is metropolitan statistics areas where more wealth in the U.S. is concentrated. Thus, the results could only account for the phenomenon in wealthier locations. Also, simply counting the number of crimes in this study may provide a distorted picture of how public safety varies over space because of spatial differences in the distribution of crimes and reporting behavior. This may also explain the positive sign for property crime. Comparing the magnitude of the coefficient for both types of crimes, we can find that violent crime has a larger effect than property crime, which means that people care more about the number of violent crimes. Comparing the results for 2005 and 2010 I find that the elasticity of WTP with respect to property crime changes slightly, but the elasticity with respect to violent crime increases by 36% which indicate that in 2010 people are willing to pay more to move to a location with lower violent crime occurrence. To explain what the estimates implied with more detail, we can consider the following example. In 2005, the number of violent crime in Abilene, TX is 640, while in the same year the number of violent crimes\ in Albuquerque, NM is 6,630 which is roughly ten times as big as Abilene. The estimated elasticity of WTP with respect to violent crime in 2005 is -0.11 which implies that the decrease in violent crime by moving from Albuquerque to Abilene would correspond to increase in WTP of 98%.

Now it comes to the analysis of MWTP for the decrease in crime occurrences. As mentioned before in this study, $\tilde{\beta}_C = \text{MWTP}_i \times \frac{C_j}{I_{i,j}}$, thus I can recover the MWTP for crime rate using this formula. During the calculation, I used the median value of household income and both kinds of crime in the full sample, which measured the median household's WTP for the decrease in crime. In 2005, the median value of household income is $71,000 (in 2005 dollars) and the median number for violent and property crimes is 12 and 93[7] respectively. In 2010, the median household income is $71,651 (in 2005 dollars)[8] and the median number for both kinds of crime is 11 and 105 respectively. Applying the formula, the calculated MWTP for a one hundred unit decrease in violent crime is $651 and $977 for 2005 and 2010 respectively. MWTP for a one hundred unit increase in property crime is $23 and $27 for 2005 and 2010 respectively. From the MWTP we can also conclude that people care more about violent crime and are willing to pay more money to move to a safer place. However, though people still willing to pay more money to move to a place where the property crime occurred more frequently, the amount they are

[7] The number of crime is measured by hundred occurrence.
[8] The CPI inflation calculator is used to transform the 2010 dollars into 2005 dollars which will facilitate the comparison.

willing to pay is small. The WTP elasticities and MWTP for property crime and violent crime are shown in table 1.3.

Table 1.3 Marginal Willingness to Pay for Property Crime and Violent Crime

	2005		2010	
	Property Crime	Violent Crime	Property Crime	Violent Crime
WTP elasticity	0.03	-0.11	0.04	-0.15
MWTP	$23	$651	$27	$977

Note: MWTP is calculated by multiplying the WTP elasticity by the median household income and dividing by the median number of each type of crime. The median household income is 71,000 and 71,651 (in 2005 dollars) for 2005 and 2010 respectively. The median number of property crime and violent crime in 2005 is 93 and 12, and the median number of property crime and violent crime in 2010 is 105 and 11. All the crimes are measured by hundred occurrences.

The estimated coefficients for other location attributes include the fraction of population employed and per capital personal income which may reflect the economic level of the location. Both the variables are statistically significant. Though we expect that the fraction of population employed may affect household location choice in a positive way, the results show people are willing to pay less to move to a location with higher fraction of population employed. This may be explained as the fact that, if the fraction of population employed is high, it means that more work place has been taken and there is less possibility that a person will find a job in this location. The results also show that metropolitan areas with higher per capital personal income are more appealing for people.

5.3 Results of Difference in Difference Analysis

The results for difference in difference analysis are shown in table 1.4. As shown before, there are two treatments here: one is the increase in police number (D_1) and the other is the decrease in crime rate (D_2). The police variable is measured by the change of the total number of police employed in each MSA. Our interest in this estimation is the coefficients for all the three terms interacted with time period T (D_1*T, D_2*T and D_1*D_2*T). For the sorting model, the coefficient for D_{1t} is -0.39 and statistically significant at the 10% level. This means that people are willing to pay 39% less to move to a location with higher numbers of police. This can be explained by the fact that the increase in police number may be an indicator of high crime rate and also people need to pay more by tax to cover the cost of additional police. The effect of the crime rate decrease on WTP is not statistically significant in the sorting model. However, people are willing to pay 41% more to move to a location in which the crime rate decreases and the police number increases. Though there are maybe more police in bad areas generally, when controlled by numbers of crimes then more police is a good thing.

When comparing the coefficients estimated from sorting model with those get from the hedonic price model, it should be pointed out that the WTP elasticities are not directly comparable. The coefficients from the hedonic price model represents the change of house prices associated with the decrease in crime rate and increase in police force while the estimates from the sorting model not only reflect the change in house prices but also the change in income and disutility from moving. Thus, the estimates from hedonic price model may be misleading. The coefficients in the hedonic price model represent house prices elassticities with respect to the two treatments. To translate the house prices elassticities into MTP, we need to first multiply the coefficient from the hedonic price model by the share of income spent on house, and here I use the average share of 2005 and 2010 which is 0.32 to calculate MTP. Thus, the elasticity of WTP for the police number increase is 0.35 which is slightly lower than that from the sorting model. However, in the hedonic price model, the true effect of the crime rate decrease and the effects from both treatments are not statistically significant. For the other location characteristics, people are willingness to pay less to move to a location where the employment rate is high and metropolitan areas with higher per capital personal income is more appealing for people.

Table 1.4 Results for Difference in Difference Estimation

Variable	Description	Sorting Model	Hedonic Model
T	Time period	-0.78*	0.42
D_1	=1 if police number increase	0.39**	1.09**
D_2	=1 if crime rate decrease	0.05	0.13
D_1*T	The true effect of police number increase	-0.39*	-1.08*
D_2*T	The true effect of crime rate decrease	-0.22	-0.72
D_1*D_2*T	The true effect of both police number increase and crime rate decrease	0.41*	1.15
Log(Employment)	Fraction of population employed	-0.65*	-1.63
Log(PerInc)	Per capital personal income	0.46**	1.21**

Note: *** means statistically significant at 1% or above; ** means statistically significant at 5% level and * means statistically significant at 10%. Crime data are obtained from FBI crime report of 2005 and 2010. Income and employment data are obtained from Bureau of Economic Analysis of 2005 and 2010.

6. Conclusions

In this study, I estimate the effect of crime on household location choice using a two stage residential sorting model where the choice set is defined at the level of the metropolitan areas. In the first stage a discrete choice model is estimated to get the MSA level fixed effect and in the second stage, these fixed effects are estimated on the number of property and violent crime and other location attritions. In this study, the household head is regarded as the decision maker and the characteristics of the household head are used to predict their income in each MAS. In order to get the house prices, a linear function is used to regress house value on a set of dwelling attributes. Finally, a difference in difference model is introduced to analyze the effects of crime rate decrease and police force increase on households' WTP.

The first stage estimation results show that living out of individuals' birth states and birth regions have negative effect on their utility and cost continue to increase with living out of one's birth region and macro region. Also, the results for short-run mobility cost show that living in a different state than one year before has a significant utility cost in both 2005 and 2010 and the cost continues to increase with living in a different region than one year before. However, the cost of living in a different macro region than one year before is not statistically significant in 2010. This may be explained by the fact that in the short-run there are fewer people changing the macro region where they live. The focus in the second stage analysis is the estimated coefficients on the number of violent and property crime which represent the elasticities of WTP with respect to these two kinds of crime. The results show that people are willing to pay more to move to a location with lower violent crime occurrence and are also willing to pay more to move to a place with higher property crime. This can be explained by the fact that higher house prices are more enticing to property crime because the value of the goods inside is expected to be more than that of a lower priced home. Also, people with large amounts of money can spend more on security devices for their home so that non-violent crime does not deter them. When recovering the WTP for the two types of crime using elasticities, it shows that people are willing to pay $651 and $977 for a one hundred unit decrease in violent crime and $23 and $27 for a one hundred unit increase in property crime for 2005 and 2010 respectively, which indicates that, though people still willing to pay money to move to a place where the property crime occurred more frequently, the amount they are willing to pay is small. The difference in difference results for the sorting model show that people are willing to pay less to move to a location in which the police number increases and pay more to move to a location where the crime rate decreases and police force increased. This can be explained by the fact that the increase in police number may be an indicator of high crime rate as well as higher tax payments to cover expenditures on additional police. When comparing the difference in difference results for the sorting model with that for the hedonic price model I find that the elasticity of WTP for the police number increase in the hedonic price model are slightly lower than that from the sorting model.

Appendix Tables:

Table A1 Description of Census Variables

Variable	Label	Mean 2005	Mean 2010
ACRE10	House on 10 acres or more	0.024	0.022
ROOMS2	2 rooms in dwelling	0.003	0.003
ROOMS3	3 rooms in dwelling	0.017	0.014
ROOMS4	4 rooms in dwelling	0.064	0.062
ROOMS5	5 rooms in dwelling	0.184	0.167
ROOMS6	6 rooms in dwelling	0.233	0.211
ROOMS7	7 rooms in dwelling	0.186	0.179
ROOMS8	8 rooms in dwelling	0.144	0.146

ROOMS9	9 rooms in dwelling	0.169	0.215
BEDROOMS1	1 bedroom in dwelling	0.014	0.015
BEDROOMS2	2 bedrooms in dwelling	0.135	0.129
BEDROOMS3	3 bedrooms in dwelling	0.479	0.466
BEDROOMS4	4 bedrooms in dwelling	0.288	0.293
BEDROOMS5	5 bedrooms in dwelling	0.083	0.095
BUILDYR1	Less than 10 years old dwelling	0.114	0.175
BUILDYR 2	20-30 years old dwelling	0.169	0.151
BUILDYR 3	30-40 years old dwelling	0.148	0.131
BUILDYR 4	40-50 years old dwelling	0.150	0.139
BUILDYR 5	50-60 years old dwelling	0.116	0.109
BUILDYR 6	60-70 years old dwelling	0.185	0.184
BUILDYR 7	More than 70 years old dwelling	0.119	0.111
UNITSSTR1	Mobile home or trailer	0.042	0.038
UNITSSTR2	Boat, tent, van, other	0.000	0.000
UNITSSTR3	1-family house, detached	0.855	0.852
UNITSSTR4	1-family house, attached	0.056	0.061
UNITSSTR5	2-family building	0.014	0.014
UNITSSTR6	3-4 family building	0.008	0.008
UNITSSTR7	5-9 family building	0.006	0.006
UNITSSTR8	10-19 family building	0.005	0.006
UNITSSTR9	20-49 family building	0.005	0.004
UNITSSTR10	50+ family building	0.008	0.010

VALUEH	House value	293624.160	299374.700
WHITE	Race of the household head =white	0.639	0.607
MALE	Sex of the household head =male	0.230	0.265
AGE	Age of the household head>60	0.814	0.806
HSDROP	High school drop out	0.048	0.040
SOMECOLL	Complete some college study	0.220	0.223
COLLGRAD	College graduate	0.365	0.397
INCTOT_HEAD	Total personal income of household head	55369.050	60187.180
METAREA	Identification number of metropolitan statistical area		
BPL	Birth state of the household head		

Note: Because of top coding, a dummy variable is created for each number of rooms and bedrooms. Data are obtained from 2005 and 2010 American Community Survey data.

Table A2. Summary of Income regress

		2005		2010	
Variable	Description	Mean	Std Dev	Mean	Std Dev
constant	Intercept	9.683	0.665	2.880	0.731
MALE	Sex of the household head =male	0.640	0.474	0.584	0.456
AGE	Age of the household head>60	-0.231	0.471	-0.270	0.480
WHITE	Race of the household head =white	0.146	0.483	0.151	0.601
HSDROP	High school drop out	-0.277	0.777	-0.182	0.593
SOMECOLL	Complete some college study	0.215	0.551	0.278	0.520
COLLGRAD	College graduate	0.619	0.516	0.681	0.435

Note: The data used for predicting income are obtained from 2005 and 2010 American Community Survey data which can be download from IPUMS website.

Table A3. Housing Service Estimated Parameters

		2005		2010	
Variable	Description	Parameter	t Value	Parameter	t Value
INTERCEPT	Intercept	2.52	11.90	3.44	32.56
ROOMS2	2 rooms in dwelling	0.64	2.84	-0.34	-2.46
ROOMS3	3 rooms in dwelling	0.55	2.58	-0.30	-2.60
ROOMS4	4 rooms in dwelling	0.47	2.19	-0.36	-3.15
ROOMS5	5 rooms in dwelling	0.58	2.67	-0.32	-2.84
ROOMS6	6 rooms in dwelling	0.72	3.32	-0.25	-2.20
ROOMS7	7 rooms in dwelling	0.83	3.82	-0.13	-1.16
ROOMS8	8 rooms in dwelling	0.95	4.40	-0.01	-0.09
ROOMS9	9 rooms in dwelling	1.15	5.30	0.17	1.51
BEDROOMS2	2 bedrooms in dwelling	0.17	3.43	0.15	2.94
BEDROOMS3	3 bedrooms in dwelling	0.26	5.05	0.30	5.78
BEDROOMS4	4 bedrooms in dwelling	0.43	8.32	0.50	9.52
BEDROOMS5	5 bedrooms in dwelling	0.59	10.76	0.73	13.25
BUILDYR1	Less than 10 years old dwelling	0.38	19.97	0.19	9.58
BUILDYR r2	20-30 years old dwelling	0.32	17.91	0.15	7.63
BUILDYR 3	30-40 years old dwelling	0.22	12.48	0.08	4.01
BUILDYR 4	40-50 years old dwelling	0.08	4.58	-0.01	-0.67
BUILDYR 5	50-60 years old dwelling	0.12	6.44	0.00	0.16
BUILDYR 6	60-70 years old dwelling	0.11	6.55	0.02	1.00
ACRE10	House on 10 acres or more	0.23	7.66	0.16	4.70

UNITSSTR2	Boat, tent, van, other	-0.21	-0.95	0.06	0.28
UNITSSTR3	1-family house, detached	1.56	65.66	1.59	58.90
UNITSSTR4	1-family house, attached	1.70	57.25	1.72	52.87
UNITSSTR5	2-family building	1.86	41.00	1.96	38.94
UNITSSTR6	3-4 family building	1.93	34.03	1.89	31.52
UNITSSTR7	5-9 family building	1.77	28.55	1.85	26.14
UNITSSTR8	10-19 family building	1.81	26.41	1.84	26.92
UNITSSTR9	20-49 family building	1.99	28.39	1.98	24.46
UNITSSTR10	50+ family building	2.14	38.80	2.12	37.54

Note: All the data used in this estimation are obtained form 2005 and 2010 American Community Survey data.

Chapter 2 . Exposure to Environmental Risk and Neighborhood Racial Segregation

1. Introduction

When making residential location choice, household tastes vary according to both household's own characteristics and choice attributes. By making tradeoffs regarding housing and neighborhood attributes, households of different characteristics may make different decisions on the living location, which lead households in a given neighborhood to be different. Household preferences for neighbors shape the way that households sort in the housing market, and influence the level of residential segregation. Racial segregation arises if households prefer to live close to households of same race and live separately from households with different race. After Tiebout (1956) seminal paper, empirical Tiebout sorting model became one of the important tools to analyze the relationship between location choice and household preferences for local public goods. The basic idea of Tiebout sorting is that households face a large number of communities offering different level of local public goods. As households sorted to choose their most preferred community, they reveal their demand for the public goods. By applying sorting model, this study investigates the impact of environmental risk on household residential location choice in Franklin County, Ohio State. However, social interaction also plays an important role in household residential location choice (Bayer and McMillan 2012), and households are more likely to live close to households that are similar to themselves. As a result, this study also analyzes the racial segregation pattern in Franklin County and how the change of household preferences heterogeneity affects the existing racial segregation pattern.

Environmental risk in this study is measured by toxic releases from facilities in the neighborhood. According to the requirement of the Emergency and Community Right-to-Know Act, facilities in different industry sectors that emit more than a given threshold level of any listed toxic substance of the Toxic Release Inventory (TRI) must report annually of the emission data to the U.S. Environmental Protection Agency's (EPA). The information about facilities that submit release level to EPA are publicly available, and the existence of this kind of facilities in the neighborhood is regarded as disamenity because of visually unattraction or unpleasant odor which may pose a threat to human health. Therefore, in this study, I analyze how the existence TRI reported facilities in the neighborhood in the neighborhood affect household residential location choice. Following Banzhaf and Walsh (2008), I also incorporate difference in difference analysis to the sorting model to identify the effect of entry of facilities that are required to report their use of chemicals to TRI on households residential location choice. Employing a two-stage equilibrium sorting model our results confirm that heterogeneous preferences exist in the process of residential location choice and entry of new TRI facilities to the neighborhood decreases mean utility significantly.

When making a residential location decision, besides choice attributes, which include housing and neighborhood characteristics, household preference heterogeneity also plays an important role, which influences the segregation level in the neighborhood. Environmental inequity has been interest of researchers for a long time, and many researches find that minorities face disproportionate exposure to various environmental risks (Hite 2000,Crowder and Downey 2010). Compared with the current neighborhood racial segregation pattern, this study applies counterfactual simulations to analyze how the racial segregation pattern changes if households' estimated heterogeneous disutility of environmental risk with respect to race is turned off. Our result indicates that differential preferences for environmental risk by race serve to integrate white households while segregate non-white households.

The remainder of this study is organized as follows. The next part is literature review, which is followed by the introduction of the equilibrium sorting model. Then, I describe the source of data and the technic used to create variables for each community. Finally, I discuss the estimation results of the sorting model and the final section is the conclusion.

2. Literature Review

Following the seminal work of McFadden (1973, 1978), many researchers have used a discrete choice framework to study residential location decisions. This section discusses previous literatures in which sorting model is used to investigate the marginal willingness to pay for the improvement of neighborhood attributes. Sorting model presented in the previous literatures always focused on the choice of housing among a set of alternatives by households and these alternatives are defined often by house characteristics and neighborhood attributes. Bayer et al. (2007) develops a framework for estimating household preferences for school and neighborhood attributes in the presence of sorting using restricted access Census data from a large metropolitan area. This study introduces a boundary discontinuity design to a heterogeneous residential choice model, addressing the endogeneity of school and neighborhood characteristics. Using a rich dataset spanning 17 years of home sales in the Twin Cities area of Minnesota, Allen Klaiber and Phaneuf (2010) analyzes how open space amenities affect residential location choices applying a horizontal sorting model. Because time in the context of sorting models is an often overlooked element, in order to capture the variation across space and time, the authors define housing type by location, house size and time. From the sorting model estimates, this study finds that heterogeneity across types of open space and across households is shown to be a critical determinant of the welfare impacts of open space conservation. Discrete sorting model is also used to investigate the impact of environmental quality on household residential location choice. Bayer and McMillan (2012) quantified the separate effects of employment geography and preference for housing attribute on neighborhood stratification using an equilibrium sorting model, and simulations based on the model and credible preference estimated show that counterfactual reductions in commuting costs lead to increases in racial segregation. The equilibrium sorting model used in our study closely follows the one used by Bayer and McMillan (2012).

Sorting models are also widely used by researchers to analyze impacts of environmental quality on residential location choice. Bayer et al. (2009) develops a discrete choice model by incorporating moving cost into the model and apply the method to the case of air quality. This study focuses on metropolitan areas throughout the US for the years 1990 and 2000 and one of the novelties of this study is that it uses pollution from distant source as the instrumental variables to deal with the correlation between air pollution and the unobservable local characteristics. However, the choice set in this study is metropolitan areas which are large areas and different in size. The great range in size of the metropolitan area may bias estimation results. For example, air pollution of a region may have effect very locally, but if it is averaged at metropolitan area, the pollution will be dilute and this may induce the correlation between air pollution with larger geographic area (Banzhaf and Walsh 2008). Tra (2010) develops a discrete choice locational equilibrium model to evaluate the benefits of the air quality improvements that occurred in the Los Angeles area following the 1990 Clean Air Act Amendments. One limitation of this empirical framework is that it focuses on the Census Public Use Microdata Sample which characterized the household residential location by a Public Use Microdata Area. However, the aggregated household characteristics at the Census Public Use Microdata Sample level provide insufficient variation.

In this chapter I analyze the impact of environmental risk on household residential location choice, in which environmental risk is measured by toxic chemical releases from TRI reported facilities. EPA's TRI data are used by many researchers to investigate its relationship to the housing market. Bui and Mayer (2003) analyze impact of toxic releases changes on house price, and find that house prices show no significant impact of declines in reported toxic releases from 1987 to 1992. Decker, Nielsen and Sindt (2005) use a cross-sectional hedonic price model to investigate the relationship between TRI data releases and the prices of single-family residence within postal zip code areas. After controlling for socioeconomic variables, they find that TRI pollutant releases affect housing values significantly. The more recent study of Mastromonaco (2015) find that listing an existing firm in the TRI lowers housing prices up to 11% within approximately 1 mile using a difference-in-difference specification. Different from the listed studies, this study uses a two-stage equilibrium sorting model to estimate household marginal willingness to pay for the change of toxic emission level, which is also reflected in housing price.

Environmental inequity has been interest of researchers for a long time, and many researches find that minorities face disproportionate exposure to various environmental risks. Hite (2000) analyzed location choice of individuals based on observed housing transactions using a random utility model and found that African American households are unfairly exposed to environmental disamenities. Crowder and Downey (2010) combined longitudinal individual level data with neighborhood-level industrial hazard data to examine the extent and sources of environmental inequality, and the results indicate that black and Latino householders move into neighborhoods with significantly higher hazard levels than whites. Therefore, the other focus of this study is the relationship between racial segregation and household heterogeneous preference for environmental quality.

3. Empirical Methodology

3.1 Conceptual Model

The sorting model used in this study follows closely the one developed by Bayer et al. (2007, 2012). Sorting model begins with a simple assumption that the amount and characteristics of houses and public goods varies across locations, and household chooses a particular housing type to maximize utility. Before proceeding to the model, it is important to define the housing type. The type of housing is characterized by both the characteristics of the house (e.g., age, number of bathrooms and so on) and its neighborhood attributes (e.g., sociodemographic composition, environmental quality, community poverty level and so on). In this study, I characterize the households' residential location choice alternatives in terms of housing units, which means that each house in our dataset represent one housing type. Each household choose the dwelling location h from a set of housing types H. Let X_h represent the observable characteristics of housing type h and let ρ_h denote its price. Then, the explicit indirect utility function form is defined as:

$$(1) \qquad V_{ih} = \beta_{iX}X_h - \beta_{i\rho}\rho_h + \xi_h + \varepsilon_{ih}$$

where V_{ih} represents the indirect utility of household i's from choosing housing type h , which is composed of the observed characteristics of the house X_h, housing price p_h, unobserved attributes of the housing type ξ_h and the idiosyncratic error term ε_h^i. Households select the housing type which provides them the highest utility. $\beta_{i\rho}$ and β_{iX} are parameters to be estimated, The heterogeneous preference of the households can be expressed by the interaction of housing characteristics and housing price with observed households' characteristics. As a result, the function of the household $i's$ taste for attribute k is given by:

$$\beta_{ik} = \beta_{0k} + \sum_{q} z_{iq}\beta_{qk} \tag{2}$$

where q is the number of household characteristics and z_{iq}denotes the q_{th} characteristic of household i. This expression decomposes household's taste for attribute k into two parts: β_{0k} captures the part that is common across all households and β_{qk} represents the taste that varies according to household's observed characteristics.

To get the locational equilibrium, I assume that the idiosyncratic error term ε_{ih} is identically and independently distributed and has a Type I Extreme Value distribution. Given this assumption, the probability household i select housing type h can be calculated and I denote it as Pr_{ih}. Using these probabilities, the predicted aggregate demand for housing type h is obtained by integrating the choice probabilities over the sample population:

$$D_h = \sum_{i} Pr_{ih}(X, Z, \rho) \tag{3}$$

The market clearing condition implies that the demand for houses of type h must equal to the supple of such houses, and then we have:

$$S_h = D_h = \sum_{i} Pr_{ih}(X, Z, \rho) \tag{4}$$

Given the indirect utility defined in equation (1) and a fixed set of housing and neighborhood attributes, Bayer et al. (2004) shows that a unique set of prices clears the market.

3.2 Econometric Implementation

The econometric model identifies the parameters defined in equation (1) and (2). A two-step strategy is used in this study. In the first step I estimate household preference parameters and the alternative-specific tastes while in the second step the mean taste parameters are recovered. Before proceeding to the two-step estimation strategy, we need first substitute equation (2) into equation (1) and rewrite equation (1) as:

(5) $$V_{ih} = \delta_h + \theta_{ih} + \varepsilon_{ih}$$

where

(6) $$\delta_h = \beta_{0X} X_h - \beta_{0\rho} \rho_h + \xi_h$$

and

(7) $$\theta_{ih} = (\sum_q z_{iq} \beta_{qX}) X_h - (\sum_q z_{iq} \beta_{q\rho}) \rho_h$$

δ_h defines variables common to all the households regardless their characteristics and θ_{ih} defines variables unique to households which arise from differences in the observed characteristics of household.

With this expression of the utility function, the first stage estimation recovers the alternative-specific constant, δ_h, and the household-specific taste parameter in θ via maximum likelihood estimation (MLE). With any combination of heterogeneous parameters in θ and the alternative-specific constant, δ_h, the probability that each household i chooses house type h can be predicted. In order to estimate the first stage, it is assumed that ε_{ih} is identically and independently distributed and has a Type I Extreme Value distribution. Then the conditional logit probability of household i choosing housing type h is defined as:

(8) $$Pr_{ih} = \frac{e^{\delta_h + \theta_{ih}}}{\sum_m e^{\delta_m + \theta_{im}}}$$

The log-likelihood for the household choices is defined as:

$$(9) \qquad \ell\ell = \sum_{i}\sum_{h} \mathrm{I}_{ih} \ln(Pr_{ih})$$

where I_{ih} is a dummy variable that equals 1 if household i chooses housing type h and 0 otherwise. The first stage estimation procedure aims to search parameters of θ and the vector of mean indirect utilities δ to maximize $\ell\ell$.

However, when the choice set is large like in this study, the estimation of the conditional Logit model is computational restrictive. Because of the independence of irrelevant alternatives assumption of the logit specification, following McFadden (1978), the estimation can be simplified by using a subset of non-chosen alternative for each household along with the chosen alternative to get the household specific taste parameters. In order to get the mean taste parameters, of which there is one for each housing type, a contraction mapping, which is proposed by Berry (1994), is used. The mean taste parameters got from this method is consistent with the maximum likelihood.

In the second stage, the mean indirect utility get from the first stage is decomposed into observable and unobservable components as shown in equation (5). The strongest prediction of the sorting model is that the entrance of TRI facility should make households to move out of the community leading to the decrease in house demand, which will decrease house price. To test this hypothesis, this study incorporates difference in difference analysis to the second stage estimation, and regresses the alternative-specified fixed utility δ_h on the entry of facilities that are required to report their use of chemicals for TRI:

$$(10) \qquad \delta_h = \beta_{0X} X_h - \beta_{0\rho}\rho_h + \alpha_I I_h^{new} + \alpha_T T_h^{post} + \alpha_{IT} I_h^{new} * T_h^{post} + \xi_h$$

where δ_h denotes mean indirect utility for housing type h; I_h^{exit} is an indicator variable for whether the household located in a community that went from no toxic exposure to some exposure from 2000 to 2010; T_h^{post} equals 1 if year equals 2010; α_I, α_T and α_{IT} are parameters to be estimated; and all the other variables are defined as before. The focus of this study are the coefficients on α_{IT}, which captures the pure treatment effect of new TRI exposure.

Before proceeding to the next part, another problem needs to be solved. The underlying assumption of the second-stage regression is that housing prices are uncorrelated with unobserved characteristics of residential locations. However, the fact is that two identical houses located in the neighborhood with identical characteristics may have different prices because that attributes of distant neighborhoods in the same housing market are likely to affect local house prices. To solve this endogeneity, following Bayer et al. (2007), this study develops an instrument for house price based on spatial nature of housing market. I assume that attributes of distant houses and neighborhoods that are located within 2 miles of the house to directly affect utility, and use exogenous attributes of distant neighborhoods as instrument for price. There are two steps to construct the instrumental variables.

The first step is to rearrange equation (10) by moving the price to the left hand side:

$$(11) \qquad \delta_h + \beta_{0\rho}\rho_h = \beta_{0X}X_h + \alpha_I I_h^{new} + \alpha_T T_h^{post} + \alpha_{IT} I_h^{new} * T_h^{post} + \xi_h$$

Then, a plausible value for $\beta_{0\rho}$ need to be guessed, which I denote it as $\widetilde{\beta_{0\rho}}$ and additional distant neighborhoods attributes N_h also needed to be added to the equation. Since in this study, the current community is created by randomly drawing one mile diameter circles in the study area, the distant neighborhoods are defined as regions that are within 1 and 2 miles away from the center of the current communities. The new equation is expressed as:

$$(12) \qquad \delta_h + \beta_{0\rho}\rho_h = \beta_{0X}X_h + \beta_{0hj}N_{hj} + \alpha_I I_h^{new} + \alpha_T T_h^{post} + \alpha_{IT} I_h^{new} * T_h^{post} + \xi_h$$

where N_{hj} denotes the j_{th} distant neighborhood attribute for housing type h. With these new variables, equation (12) is estimated using ordinary least square (OLS). By setting the OLS residual, ξ_h, to zero, the instrument for housing price is obtained as follows:

$$(13) \qquad \rho_h^{iv} = \frac{\left(\widehat{\beta_{0X}} X_h + \widehat{\beta_{0hJ}} N_{hj} + \widehat{\alpha_I} I_h^{new} + \widehat{\alpha_T} T_h^{post} + \widehat{\alpha_{IT}} I_h^{new} * T_h^{post}\right) - \delta_h}{\widetilde{\beta_{0\rho}}}$$

One problem I need to point out here is that the instrument is dependent on the initial value of $\widetilde{\beta_{0\rho}}$. In order to eliminate this dependence, I apply the method used by Klaiber (2010). The strategy is that after guessing the initial value of $\widetilde{\beta_{0\rho}}$ we can get the first ρ_h^{iv} and running the regression with this instrument for price the price coefficient is obtained. Then I replace this coefficient with the initial one and estimation the equation with the second value of ρ_h^{iv}. This procedure is repeated until the price coefficient stabilizes.

4. Data Sources

4.1 Definition of Communities

In order to get the effect of entry of new TRI reported facility in the neighborhood on residential location choice using difference in difference analysis, it is required that the boundaries of the communities in both 2000 and 2010 remained fixed. In previous study (see Allen Klaiber and Phaneuf 2010, Tra 2013) census block groups or census tract are used as community. However, the boundaries of many census block groups had changed between decennial censuses. Another problem of using census tract as the community is that the sizes of the communities are different and the quality of public goods may be lower when averaged over a large area, which may bias the results. For these reasons, Banzhaf and Walsh (2008) use a different approach for neighborhood definitions and define neighborhoods as a set of half-mile-diameter circles evenly distributed across the study area. In this study, I use the same method to define the neighborhoods. The communities are constructed by placing an equidistant grid across the Franklin County of Ohio State. Both the width and height of the grids is 0.5 mile. After the grids have been constructed, a 0.25 mile buffer is placed inside each grid, creating a set of circles of 0.25-mile radius (alternatively 0.5-mile-radius circles) that are evenly distributed across the study area. This process creates 2481 “0.25-mile radius communities” and 542 “0.5-mile radius communities” in the study area. Figure 2.1 shows the approach used to assign demographic data to the new communities across the study area with the help of the ArcGIS software. In the figure, the circles are the new communities and the demographic data are assigned to communities based on the percentage of the census block’s geographic area lying within each community. Taking population for example, 49.13% of the population in census block “390490094204003” and 84.82% of the population in census block “390490093262012” is assigned to the selected community in figure 2. Similarly all other demographic data from the 2000 and 2010 census are aggregated to the new circle-communities. Table 2.1 shows the descriptive statistics of the demographic variables for the “0.25-mile radius communities” in 2000 and 2010. The demographic characteristics for each community are easily comparable because of the same size (approximately 0.1963 square miles). The mean population of all the communities is 384 and 420 for 2000 and 2010 respectively, which means that total population of Franklin County increased during the study period. However, the percentage of white population decreased while percentage of black population and other increased. This indicates that the increase in total population is caused by the increase of black and other race population. In my study, I define these 0.25-mile radius communities as the neighborhood.

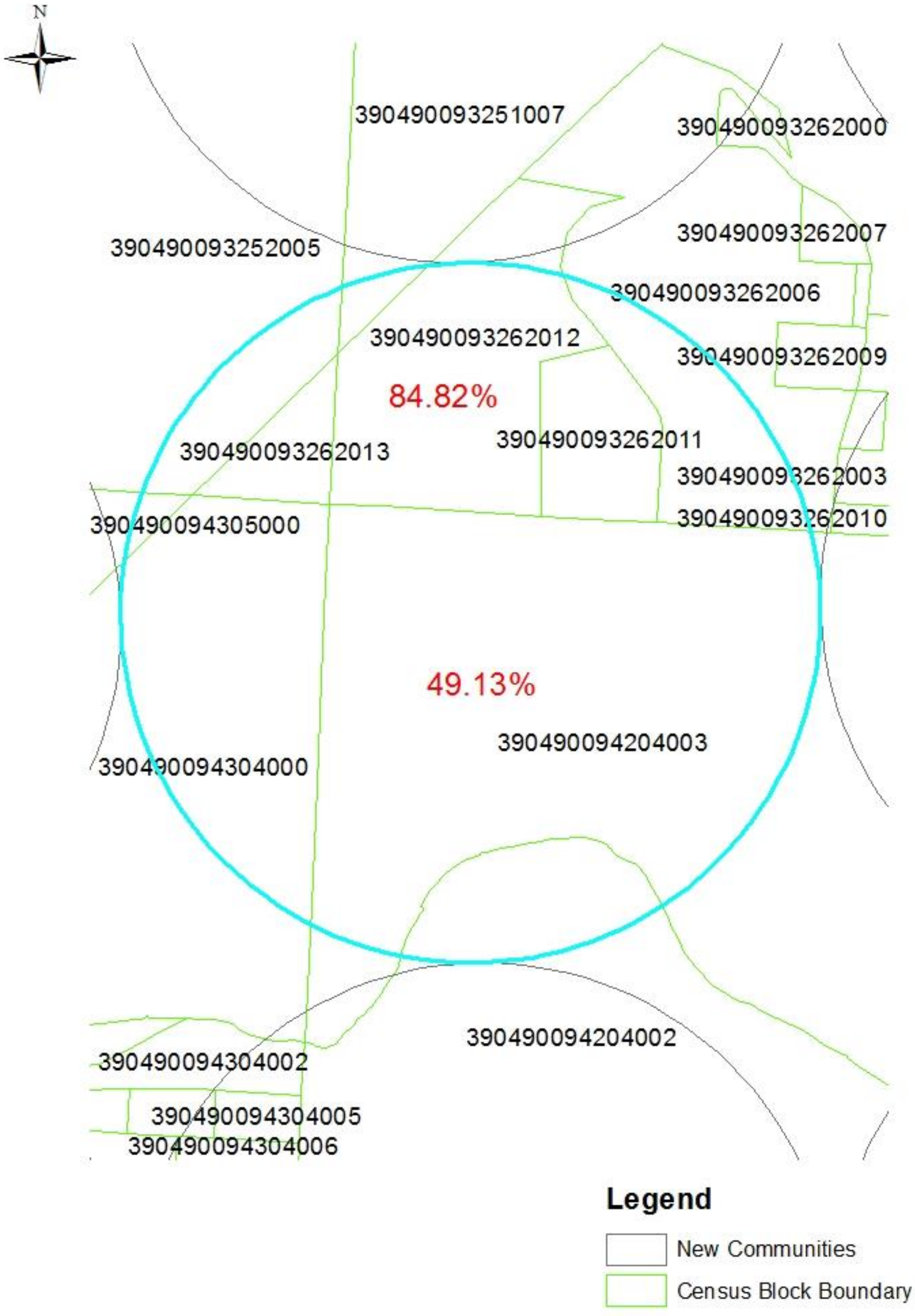

Figure 2.1 New Communities Creation

4.2 Household and Housing Characteristics

The housing data include residential real estate transaction data of Franklin County in Ohio State, and I only focus on single-family residential properties transacted in 2000 and 2010. The property transaction data provide transaction records for residential properties located in Franklin County. Each record includes the property's address, transaction price and the structure characteristics of the house. Each transaction is assigned to a community based on its geographic location. After cleaning the data, there are 10855 housing transaction records in 2000 and 12955 transaction records in 2010, which means that there are 10855 housing types in 2000 and 12955 housing types in 2010.

Though housing transaction data provides precise location for each house and a comprehensive set of housing characteristics, these data do not provide information about the households occupying the houses. As a result, aggregate information need to be used to approximate household characteristics. In this study, household characteristics and neighborhood characteristics are approximated using block and block group level data from 2000 and 2010 census data, which are publicly available. The household level variables which are approximated at census block level include household size and householder's race[9], household income is approximated at census block group level[10]. Though there is no variability in households selecting living in the same community, but there is variability among the full set of households across the whole study area. The mean value of structure characteristics and house prices of all

[9] To assign the race to each householder, I first calculate the composition of the race for each census block, if the percentage of white population in the block is higher than the average value of the whole sample, and then I assume that all the householders in this block are white. Same logic is used to define black household and household of other race.
[10] Household income data for each block group in 2010 are obtained from 2007-2011 American Community Survey 5-year estimate.

housing types are shown in table 2.1. The mean house price had increased by 6.5% from 2000 to 2010. Except house age, other structure characteristics of the houses did not change too much during the study period. Household characteristics include household race, household income and household size. In this study, I classify household race into three categories: white, black and other. From 2000 to 2010, average household income increased by almost 20%, and average household size is about 2.5 in both 2000 and 2010. Regarding the race of householder, percentage of houses occupied by white decreased by 4%, while percentage of non-white households increased by 4%.

Table 2.1 Summary Statistics for Key Variables

	2000		2010	
	Mean	Standard deviation	Mean	Standard deviation
Neighborhood characteristics				
Population	384.47	457.04	419.71	438.53
Population density	1958.61	2328.28	2138.11	2233.97
Percent White	0.77	0.30	0.73	0.30
Percent Black	0.17	0.28	0.19	0.27
Percent other races	0.06	0.11	0.08	0.13
Poverty level	0.08	0.09	0.12	0.12
Release	870.50	10414.62	367.54	5252.47
Number of observations		2481		2481
Housing characteristics				
Price	142861.28	77601.48	152163.74	107035.28

Bathrooms	1.95	0.71	1.87	0.75
Fire place	0.56	0.59	0.52	0.61
Air conditioner	0.88	0.32	0.81	0.39
House age	34.41	112.97	49.83	103.29
Number of observations		10855		12955
Household characteristics				
White	0.69	0.46	0.65	0.48
Black	0.25	0.43	0.28	0.45
Other	0.06	0.23	0.08	0.26
Income	48371.44	22897.27	57948.41	31933.42
Household Size	2.49	0.70	2.47	0.73
Number of observations		10855		12955

4.3 Neighborhood Attributes

The purpose of this study is to estimate the effect of toxic chemical release on households' residential location choice and how the entry of new TRI reported facilities in the neighborhood influences households' residential location decision. Toxic chemical releases from the facilities throughout the Franklin County are provided by EPA's TRI data set. The TRI is a publicly used database that contains information on toxic chemical releases and other waste management activities reported annually by facilities in certain industrial sectors. The reports contain information about the types and amounts of toxic chemicals that are released each year to the air, water, land and by underground injection. Most importantly, the EPA also gives the latitude and longitude of each facility and this geographic information enable us to combine the environmental quality data to the communities. When using TRI data, there are some problems: 1) EPA only requires facilities that use or produce more than threshold amount of listed toxic substances to report releases, and facilities may go in and out of reporting even if they are continually emitting toxic chemicals; 2) the list of toxic substances has expanded over time to include more industries and chemicals. To reduce the misleading caused by these problems, I keep every facility that ever reported toxic emissions to the EPA between 1997 and 1999 for year 2000 and every facility that ever reported toxic emissions between 2007 and 2009 for year 2010. I also use the 3-year lagged average toxic emission amount for each facility to represent the release level. Figure 2.2 shows the distribution of the TRI facilities in the study area, and there are 85 TRI reported facilities in 2000 and 67 facilities in 2010.

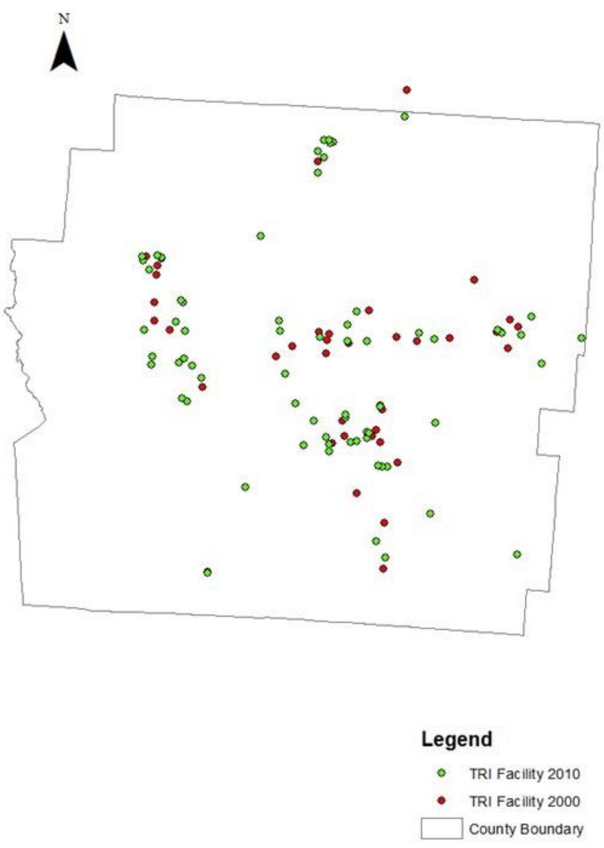

Figure 2.2 TRI Facility Distribution

Method used to assign release level to each community is the same as that used to assign demographic characteristics to each community. In this study, I assume that toxic chemicals released from the TRI facilities only affect communities within 0.25 mile (alternatively 0.5 mile) from the facility. To assign the release level for each community, I first draw a 0.25-mile radius circle for each facility to represent the pollutant area, and then the release levels are assigned to communities based on the percentage of the pollutant geographic area lying within each community. There are 196 and 147 communities exposed to toxic release in 2000 and 2010 respectively. This analysis considered aggregate TRI releases data rather the disaggregated data for each release type (releases into water, releases into the air or releases transported offsite).

Before proceeding to the estimation results, I need to consider the limitations of our analysis. Firstly, I assume that households have full information about all the housing and neighborhood characteristics, and in this study I assume that the disamenities caused by TRI facilities are perceived by all the households. However, both informed and uninformed households exist in reality. Hite (1998)using survey data about home buyers' knowledge of nearby landfill sites found that though home buyer are poorly informed about disamenities, those who are informed bid down house price. Secondly, I arbitrarily define the new communities as a set of 0.25-mile radius circles and do not know whether the estimation results are sensitive to the community size. For this reason, I employ a sensitivity analysis by replicating the equilibrium sorting model with 0.5-mile radium communities. Similarly, I also do a sensitivity test for the pollutant area using 0.5-mile radius buffer instead of 0.25-mile radius buffer. Finally, facilities report their chemical emission for TRI only when their release amount exceeds the threshold, and therefore the TRI emissions are censored. This means that our results could not capture the effects of facilities with lower level emission.

5. Estimation Results

Using dataset creased above, I investigate the effect of toxic chemical release on households' residential location choice and how the entry of new TRI reported facilities in the neighborhood influences households' residential location decision. Since households of different race may face disproportionate exposure to environmental risks, I also use counterfactual simulations to analyze how the racial segregation pattern changes if household preference over environmental risk exposure does not change with household race. Our primary results focus on estimations using 0.25-mile radius communities and 0.25-mile radius buffer around TRI facilities, and I use 0.5-mile radium in sensitivity analysis respectively.

5.1 First Stage Estimation Results

The model regressed in this study is at the level of house types, and all the choice alternatives are characterized by both housing and neighborhood characteristics. As described before, a two stage equilibrium sorting model is used and the purpose of the first stage is to recover the household-specific taste parameters as well as a vector of mean indirect utilities for each housing type. In specifying the model, I include a limited set of interactions between household characteristics and the neighborhood attributes. By reducing the number of interactions according to the reasonable intuition, the degree of freedom for the estimation could be conserved and the potential problem of collinearity could also be limited. The estimation results for the interactions between household characteristics and housing attributes are shown in table 2.2.

Table 2.2 First Stage Estimation Results

Parameter	Estimate	Standard Error
Household size * Number of bathrooms	0.022***	0.005
Black * Poverty level	0.242*	0.129
Other race * Poverty level	-0.314	0.276
Black * Percent black	3.660***	0.047
Other race * Percent other race	8.976***	0.254
Black * Release	0.076***	0.029
Other race * Release	0.090*	0.054
Income * Release	-0.014***	0.005
Black * House price	-0.009***	0.002
Other race * House price	-0.013***	0.004
Income * House price	0.001***	0.000

Note: *** indicates statistical significance at 99 percent level; ** indicates statistical significance at 95 percent level; * indicates statistical significance at 90 percent level.

The coefficients on the interaction indicates that there exits variation in the preferences of households for the characteristics of their housing choice. Rather than the magnitude of the coefficients, I am more interested in the sign. All signs for the interactions shown in the results are as expected and statistically significant except the coefficient on the interaction between poverty level and other races. Coefficients on interaction of household race, household income with TRI release level indicate households' heterogeneous preference for neighborhood environmental quality. Since I do not have the accurate information about householder, the race of householder is assigned to each household according to the demographic composition of the census block that the household resided. Results in table 2.2 indicate that non-white households are more likely to sort into neighborhood with higher toxic release level, and households with lower income are more likely to choose houses expose to higher toxic releases. In the estimation, I also include other interactions: interaction between household size and the number of bathrooms, interaction between household income and house price, and interaction of household race with neighborhood poverty level, neighborhood demographic composition and house price. The interaction between household size and the number of bathrooms takes on a positive sign which is in accordance with the hypothesis that larger households prefer houses with more bathrooms. The coefficient on the interaction of household income and house price is positive which implies that, controlling for all the other factors, increase in household income will increase the housing demand. The negative signs for coefficients on interactions between household races and house price indicate that non-white households prefer to select houses with lower price. Coefficient on interaction between black and poverty level is positive, indicating black households are more likely to sort into communities with higher poverty level. While signs of the coefficients in table 2.2 give us some idea about households' heterogeneous preference for housing and neighborhood attributes, the exact interpretation of these parameters in terms of marginal willingness to pay needs the estimation results of the second stage.

5.2 Second Stage Estimation Results

To estimate the average effect of new TRI exposure on households' residential location choice, I introduce difference in difference analysis in the second stage. Using the estimation results of the mean taste parameters from the first stage estimation as dependent variable, the second stage estimation can be implemented. As shown in equation (10), there may be correlation between housing prices and unobserved housing/neighborhood characteristics in the second stage estimation. For example, two identical houses in neighborhood of identical quality may have different prices, which depend on how they are situated compared with other houses in the nearby communities (Bayer, Ferreira and McMillan 2007). To solve this endogeneity problem, an instrumental variable is introduced in the second stage, which is created based on equation (12) to equation (14). To create the instrument, I include neighborhood variables designed to account for observable determinants of housing prices. These variables are the same as variables estimated in the second stage for the 1 and 2 mile ring around each community centroid. With this instrument in place an IV regression of equation (10) is run and the results are reported in table 2.3.

Table 2.3 Instrumental Estimation Results for Second Stage

Variable	Parameter	Standard Error
Intercept	1.055	0.087
House price	-0.094***	0.006
Bathrooms	0.447***	0.042
Fire place	0.266***	0.030
House age	-0.021*	0.013
Air conditioner	0.096**	0.042
Release (1000 pounds)	-0.570***	0.024
Time * New TRI facility	-0.247*	0.657
Time	0.306***	0.033
New TRI facility	0.231*	0.653
Percent black (0.01)	-1.368***	0.041

Percent other races (0.01)	-1.884***	0.127
Poverty level (0.01)	-1.687***	0.161
Population density (Per square mile)	-0.081***	0.023

Note: *** means statistically significant at 1%; ** means statistically significant at 5%, and * means statistically significant at 10%. Variable "New TRI facility" is a dummy variable which equals 1 if a certain community experienced from no toxic emission exposure to toxic emission exposure during the study period, and variable "Time" equals 1 if year is 2010.

Parameters estimated in the second stage returned the mean preferences for housing and neighborhood characteristics. Table 2.3 shows that the coefficient on price is negative and statistically significant, which means that houses with higher price provide lower utility. Houses with more bathrooms and equipped with fireplace and air conditioner are more preferred by households, and older houses provide lower utility. Coefficient on toxic release is negative and statistically significant, which indicates that the exposure of toxic release will decrease the mean utility provided by the house. Another focus of this study is the treatment effect of new TRI exposure, which is captured by the coefficient on the interaction between release level and time dummy. Result in table 2.3 shows that entry of new TRI facilities to the neighborhood decreases mean utility significantly. I also include some other control neighborhood characteristics in the second stage estimation which include poverty level, population density, and neighborhood demographic composition. As expected houses located in communities with higher poverty levels, higher population density and higher percentage of non-white population provide lower utility while communities.

To verify the robustness of above results, I employ some sensitivity analysis. Firstly, I tested an alternative definition of the toxic release exposure variable using 0.5-mile radius buffers around each TRI reported facility instead of 0.25-mile radius buffers. I also replicated the estimation process with 0.5-mile radius communities. Though the magnitude of the mean willingness to pay from these sensitivity analyses is lower, the qualitative nature of the results does not change.

Combining the estimation results from first and second stage, households' heterogeneous marginal willingness to pay for each housing and neighborhood characteristics could be estimated. To get the marginal willingness to pay, I first calculate the marginal effect for each observation and then take average. As shown in the first stage estimation, households have heterogeneous preference for housing and neighborhood attributes. Therefore, here I estimate heterogeneous marginal willingness to pay based on different household characteristics. The estimation results are shown in table 2.4, and I just focus on heterogeneous marginal willingness to pay for neighborhood demographic composition and environmental risk. The first row of table 2.4 reports the mean marginal willingness to pay for the change listed in the head column. The mean willingness to pay shows that if the neighborhood release level decreases by 1000 pounds, house price in this neighborhood will increase by $5321, and a 1% increase in black population and population of other race will decreases house price by $1282 and $1787 respectively. The remaining rows describe the difference in willingness to pay associated with the change listed in the heading column keeping all other factors constant. Starting with the second row of the table which describes the difference in marginal willingness to pay for a black versus white household, the result indicates that black households are willing to pay $3438 more for a 1% increase in the faction of black than white households. The third row of table 2.4 shows that households of other races would like to pay $8613 more for a 1% increase in the fraction of same race neighbors than white households. The last row indicates that with each $10000 increase in household income, a household's marginal willingness to pay increases by $591 for a decrease of 1000 pounds of releases in the neighborhood.

Table 2.4 Heterogeneous Marginal Willingness to Pay Measures

Neighborhood attributes changes	Release (-1000 pounds)	Percent black (+ 1%)	Percent other races (+1%)
Mean marginal willingness to pay	5321	-1282	-1787
Black (vs White)	-1073	3438	-
Other races (vs White)	-894	-	8613
Household income (+$10000)	591	-	-

Note: The first row of the table shows the mean marginal willingness to pay for the change of neighborhood characteristics in the heading column. The remaining rows describe the difference in willingness to pay associated with the change listed in the heading column keeping all other factors constant.

5.3 Simulation Results

To explore the relationship between racial segregation and household preferences for environmental quality, I did an equilibrium simulation using parameters estimated in the two stage equilibrium sorting model. Since minorities were found to face disproportionate exposure to various environmental risks, in the simulation I assume that household preference over environmental risk exposure does not change with household race and turned off taste parameters of the interactions between household race and neighborhood TRI release level. Before carrying out simulation, it is useful to analyze the observed neighborhood racial segregation pattern, which provides a benchmark for the counterfactual. The degree of neighborhood racial segregation is measured by exposure rate following the definition of Bayer, et al. (2004). For a particular neighborhood, I calculate the fraction of households in each of the three race categories that reside in the same neighborhood as a given household, and average these neighborhood measures over all households of a given race[11].

The procedure of simulation is the computation of a new equilibrium. Parameters in the first stage estimation describe how household preferences for housing and neighborhood attributes vary with household characteristics. During the simulation, I turned off heterogeneous disutility of environmental risk exposure with respect to race, and calculate a new set of prices that clears the market. Taking the new prices and the initial sociodemographic composition of each neighborhood, I calculate the probability that each household chooses each housing type and aggregate these choices to the neighborhood level to get the predicted demographic composition of the neighborhood[12]. Then the initial neighborhood demographic composition is replaced by the predicted demographic composition, and new equilibrium is calculated. This process is continued until neighborhood demographic characteristics converge.

The counterfactual simulation results describe the role of heterogeneous preference with respect to household race for neighborhood environmental quality in shaping the extent of neighborhood racial segregation, and the results are shown in table 2.5.For comparison, the racial composition of Franklin County is 66.9% white, 26.4% black and 6.7% other races. Panel A of table 2.5 reports the observed race exposure rates. Taking black households for example, these measures imply that black households in Franklin County live in communities that have on average 48.1% white, 47.6% black and 4.3% other races. Comparing the measured exposure rates to the racial composition of the whole sample-66.9% white, 26.4% black and 6.7% other

[11] The measures of the average exposure, $E(r_j,\ R_j)$, of households of a race j to households of race k are expresses as: $E\left(r_j,\ R_j\right) = \frac{\sum_i r_j^i R_k^i}{\sum_i r_j^i}$, where r_j^i is a dummy variable which equals 1 if household i is of race j, and R_k^i represents the fraction of households of race k in household i's neighborhood.

[12] The predicted neighborhood demographic composition is expressed in terms of the probability that each household observed in the data chooses each house type in that neighborhood. The contribution of household j made to the demographic composition of neighborhood $n(h)$ is: $Z_{n(h)}^j = \sum_{k \in n(h)} Z^j * P_k^j$.

races-there is obvious evidence that black households live in communities with approximately 1.8 times the fraction of black households than would be found if they were uniformly distributed across the study area, and most of the additional fraction of black households in communities in which black households live is offset by a decrease of white households. The remaining race exposure rates indicate that households of each race living with households of same race in proportion are higher than their proportion for the whole Franklin County. Panel B of table 2.5 shows the counterfactual exposure rates. Switching off heterogeneous disutility of environmental risk with respect to household race leads to an decrease in segregation of 11.11% for white (as measured by the over-exposure to households of the same race), and a decrease of 8.5% and 10.5% for black and other races. This indicates that differential preferences for environmental risk by race serve to segregate households.

Table 2.5 Simulation Results of Race Exposure Rates

Panel A				
	Observed race exposure rate			
Household race	Percent white	Percent black	Percent other	
White	**74.1%**	19.5%	4.4%	
Black	48.1%	**47.2%**	4.3%	
Other	63.0%	20. 8%	**16.2%**	
Panel B				
	Simulation results of race exposure rate			Percentage change in own-race
Household race	Percent white	Percent black	Percent other	“over-exposure”
White	**73.3%**	20.1%	6.6%	-11.11%

Black	50.1%	**45.8%**	4.1%	-6.7%
Other	62.7%	22.1%	**15.2%**	-10.5%
Overall	66.9%	26.4%	6.7%	

Note: Each row gives the average exposure of households whose type is relevant to the row category to neighbors in the heading column. All the rows sum to one.

6. Conclusions

Household residential location decision is made based on both household characteristics and choice attributes. This study analyzes whether or not, and to what degree local environmental risk impact household residential location choice and how the change of household heterogeneous preference for environmental quality affects racial segregation, using Franklin County of Ohio State as the study area. A two-stage equilibrium sorting model is used to accommodate public goods and household heterogeneity. To investigate treatment effect of entry of new TRI reported facilities into the neighborhood, the difference in difference analysis is incorporated into the second stage estimation.

The equilibrium sorting model used in this allows household preference for housing and neighborhood characteristics to vary with household characteristics, and first stage estimation recovers household-specific taste parameters. Our results from first stage estimation indicates that non-white households are more likely to sort into neighborhood with higher toxic release level, and households with lower income are more likely to choose houses expose to higher toxic releases. Based on the alternative-specified fixed utility from first stage estimation, the second stage estimation returns the mean willingness to pay for each housing and neighborhood characteristics. With the whole set of estimated parameters, the heterogeneous marginal willingness to pay can be calculated. To explore the relationship between racial segregation and household preferences for environmental quality, I did an equilibrium simulation using parameters estimated in the two stage equilibrium sorting model. Since minorities were found to face disproportionate exposure to various environmental risks, in the simulation I assume that household preference over environmental risk exposure does not change with household race. In counterfactual simulation, racial segregation decreased, which means that differential preferences for environmental risk by race serve to segregate households.

It is worth noting that this analysis includes limited set of neighborhood characteristics for the reason of data accessibility, and future research should include a comprehensive set of neighborhood attribute to overcome the shortcomings of the present study. Moreover, aggregate TRI releases are used in this study and I did not consider the disaggregated effect of different releases, such as releases into water, releases into the air or releases transported offsite. Finally, I assume that all households are fully informed about neighborhood disamenities caused by TRI releases, and further research should take an effort to distinguish between informed and uninformed households.

Chapter 3 . Residential Sorting and Environmental Disamenities: The Case of Landfills

1. Introduction

The study of relationship between residential location choice and environmental disamenities has captured interest of many economists (Bayer et al. 2009, Tra 2010). When making residential location choice, households make trade-offs among a series of factors, such housing characteristics, environmental externalities, social attributes, and budget constraints. During the process of making decisions about living location based on their preferences, households sort across neighborhood, and as households sort across neighborhood, their demand for public goods, which are not traded in formal market, are revealed. Understanding consumer heterogeneity can be helpful to evaluate policies targeting public goods and externalities (Kuminoff, Smith and Timmins 2013).

The purpose of this study is to estimate whether the existence of landfills as well as the type of landfills will affect households' residential location decision making. The negative externalities associated with close proximity to a landfill site are confirmed by researchers (Hite 2001, Ready 2010).The possible disamenities associated with landfills are groundwater contamination, accumulation of methane gas, and increased traffic from transportation of waste. If these local disamenities generated by landfills are perceived by the residents, these perceptions can translate into discount of property values. Three study areas are defined include three landfills in the Frankly County of Ohio State. Two of the three landfills are licensed to accept construction and demolition debris and the third one accepts municipal solid waste. Different from previous studies of landfills that use hedonic model (Hite 2001, Ready 2010), an equilibrium sorting model is employed in this study. Sorting model was first proposed by Tiebout (1956) who responded to Samuelson's paper, and after that the sorting model became one of the important tools to analyze the relationship between location choice and local public goods. Besides including information provided by an equilibrium hedonic price function, sorting models also allow households preferences to vary with household characteristics. Based on this idea, the impact that a landfill has on household residential location choice can be identified by estimating a two-stage sorting model, where in the first stage a multinomial Logit model for the household choice is estimated and the second stage is an ordinary least square estimation at the house type level. Our theoretical framework is based on the assumption of preference heterogeneity regarding environmental disamenities, which is measured by the distance to the nearest landfill. To employ the model, a full year of 2010 real estate transaction data in Franklin County were collected. 2010 census block data and 2006-2010 American Community Survey 5-year estimate census block group data were used to create a proxy of household characteristics, and the distances from landfills to each household are created from maps. Combing all these data sources, the data set includes household characteristics, housing characteristics and neighborhood attributes.

The next section of this study lists some previous studies and compares these studies with ours. Then I develop the theoretical structure necessary for estimating the impacts of landfills on households' residential location choice, which is followed by the description of housing price data and other data sources. The last section reports results of this analysis, and the last part is conclusions.

2. Literature Review

There is a vast body of literatures analyzing the relationship between house value, environmental quality and household residential location choice. In this section I discuss some of the previous studies and make a comparison of these studies with ours. I review two relevant strands of literature according to the estimation technique: hedonic price model and sorting model.

The hedonic technique has been widely used in previous studies to measure the effects of landfills on house value, however, different studies got different results. Bouvier et al. (2000) examines six landfills, which differ in size, operating status, and history of contamination. The effect of each landfill is estimated by the use of multiple regression and the results show that five of the landfills have no statistically significant effect on house values. In the remaining case, the result indicates that houses in close proximity to this landfill suffered an average loss of about six percent in value. Hite (2001) analyzed the impact of presence of landfills on nearby residential real estate prices using a hedonic price model. The author account for temporal effects by including housing transaction in areas with both open and closed landfills and control for information effects. The results suggest that closing landfills will not necessarily mitigate property-value impacts. Kinnaman (2009) used both a hedonic pricing model and a repeat-sales estimator to estimate how a landfill closure affects neighboring property values. Results of are used in the analysis. Housing data gathered before and after the closure of a solid waste landfill suggest property values increased by an estimated 10.8% with the closure of a solid waste landfill, but this estimate is not statistically significant. Also, property values continued to rise with distance from the open or closed landfill, suggesting a potential stigma effect associated with the old landfill site. Ready (2010) used a hedonic price function to estimate a region containing three landfills that differ in size and in their prominence in the landscape. The results show that the three landfills differ in their impact on nearby property values. While two of the three landfills have statistically significant negative impacts on nearby property values, the smallest, least prominent landfill does not. Though these previous studies got inconsistent conclusions, most of them find negative effects of landfills on house values. Thus, in this study I also assume that landfills may decrease the value of nearby houses and household would not like to live near landfills.

Reviewing the previous, another technique used to estimate the impacts of environmental quality on household residential location choice is the sorting model. Sieg et al. (2004) uses a discrete continuous choice model measuring the general equilibrium willingness to pay for reductions in ozone concentrations in Los Angeles Metropolitan Area, which includes parts of five counties between 1990 and 1995. Bayer, Keohane and Timmins (2009) develops a discrete choice model by incorporating moving cost into the model and apply the method to the case of air quality. This study focuses on metropolitan areas throughout the US for the year 1990 and2000. The model yields an estimated elasticity of willingness to pay with respect to air quality of 0.34-0.42, which imply that the median household would pay $149-$185 for a one-unit reduction in average ambient concentrations of particulate matter. Tra (2010) develops a discrete choice locational equilibrium model to evaluate the benefits of the air quality improvements that occurred in the Los Angeles area following the 1990 Clean Air Act Amendments. The results show that air quality improvement provided substantial general equilibrium benefits to households, and it also reveals the welfare impacts varied significantly across income groups. In this study, I also use the sorting model. From the above literatures, I can see that most of them focus on air quality, and there is no study using the sorting model to estimate the effect of landfills on residential location choice. Therefore, this may be another innovation of my study.

3. Empirical Methodology

3.1 Conceptual Model

Sorting model begins with a simple assumption that the amount and characteristics of housing and public goods varies across locations, and each household choose its preferred location to maximize their utility. The utility function specification is based on the random utility model, which includes choice-specific unobservable characteristics. The framework of this study follows closely the sorting models developed by Bayer et al. (2007), which model the residential location decision of each household as a discrete choice of a single residence. I assume that each household choose the dwelling location h from a set of housing types H. Let X_h represent the observable characteristics of housing choice h, including characteristics of the house (e.g., number of rooms, house age, whether equipped with air conditioner, et al.). Let N_h represents neighborhood attributes and ρ_h denote the price of housing choice h. Then, the explicit indirect utility function form is defined as:

(1) $$V_h^i = \alpha_X^i X_h + \alpha_N^i N_h - \alpha_\rho^i \rho_h + \xi_h + \varepsilon_h^i$$

where V_h^i represents the indirect utility of household i by choosing housing choice h , which is composed of the observed characteristics of the house X_h, neighborhood attributes N_h, housing price p_h, unobserved attributes of the housing type ξ_h and the idiosyncratic error termε_h^i. α_j^i ($j = X, N, \rho$) are parameters that need to be estimated.

The heterogeneous preference of the households is allowed to vary with its own characteristics, z^i, which can be expressed by the interaction with observed characteristics of households. As a result, the parameter associated with housing and neighborhood characteristics and price α_j^i for $j \in \{X, N, \rho\}$, varies with household $i's$ own characteristics according to:

$$(2) \qquad \alpha_j^i = \alpha_{0j} + \sum_{k=1}^{K} \alpha_{kj} z_k^i$$

Equation (2) describes household $i's$ preference for choice characteristics j. Given the household's problem described in equations (1) and (2), household i chooses housing choice h which provides the maximum utility.

3.2 Econometric Implementation

The econometric model identifies the parameters defined in equation (1) and (2). Estimation of the model follows a two-step procedure, during which the first step estimates household preference parameters and the alternative-specific tastes while in the second step the mean taste parameters are recovered. Before proceeding to the two-step estimation strategy, I rewrite the indirect utility function as:

$$(3) \qquad V_h^i = \delta_h + \lambda_h^i + \varepsilon_h^i$$

where

$$(4) \qquad \delta_h = \alpha_{0X} X_h + \alpha_{0N} N_h - \alpha_{0\rho} \rho_h + \xi_h$$

and

$$(5) \qquad \lambda_h^i = \left(\sum_{k=1}^{K} \alpha_{kX} z_k^i \right) X_h + \left(\sum_{k=1}^{K} \alpha_{kN} z_k^i \right) N_h - \left(\sum_{k=1}^{K} \alpha_{k\rho} z_k^i \right) \rho_h$$

In equation (3), δ_h represents the utility provided by the housing choice h that is common to all households, and λ_h^i captures utility that is unique to households which arise from differences in the observed characteristics of household. z^i represents household characteristics (such as household income, household size, education attainment and race) and k indexes the kth characteristic. When the household characteristics included in the model are constructed to have mean zero, δ_h is the mean indirect utility provided by housing choice h.

With this expression of the utility function, the first stage of the estimation procedure is a Maximum Likelihood Estimation (MLE), which recovers the mean utility δ_h and the household-specific taste parameters in equation (5). For any combination of the heterogeneous parameters in equation (5) and the mean indirect utilities δ_h, the first stage predicts the probability that each household i chooses house h. I assume that the idiosyncratic error term ε_h^i is identically and independently distributed and has a Type I Extreme Value distribution. Then the conditional logit probability of household i choosing housing type h is defined as:

$$(6) \qquad P_h^i = \frac{\exp(\delta_h + \lambda_h^i)}{\sum_k \exp(\delta_k + \lambda_k^i)}$$

The log-likelihood for the household choices is defined as:

$$(7) \qquad \ell = \sum_i \sum_h I_h^i \ln(P_h^i)$$

where I_h^i is a dummy variable that equals 1 if household i chooses housing type h. Then the first stage estimation procedure aims to search parameters of θ and the vector of mean indirect utilities δ to maximize $\ell\ell$. However, when the choice set is large like in this study, the estimation of the conditional Logit model is computational restrictive. Because of the independence of irrelevant alternatives assumption of the logit specification, following McFadden (1978), the estimation can be simplified by using a subset of non-chosen alternative for each household along with the chosen alternative to get the household specific taste parameters. In order to get the mean taste parameters, of which there is one for each housing type, a contraction mapping, which is proposed by Berry (1994), is used. The mean taste parameters got from this method is consistent with the maximum likelihood.

When estimating equation (4), one important underlying assumption is that housing prices are uncorrelated with unobserved characteristics of residential locations. However, there is likely significant correlation between housing prices and unobserved housing/neighborhood attributes. To solve this endogeneity, following Bayer et al. (2007), this study also introduce an instrument variable for price that is based on the exogenous attributes of distant. It is assumed that distant neighborhoods influence prices in local neighborhoods but the characteristics of those distant neighborhoods are unlikely to be correlated with local unobservable components of utility. There are two steps to construct the instrumental variables.

The first step is to rearrange equation (4) by moving the price to the left hand side:

(8) $$\delta_h + \alpha_{0\rho}\rho_h = \alpha_{0X}X_h + \alpha_{0N}N_j + \xi_h$$

Then, a plausible value for $\alpha_{0\rho}$ need to be guessed, which I denote it as $\widehat{\alpha_{0\rho}}$ and add additional regressors to the right hand side based on the observed neighborhoods attributes and neighborhood social demographics for all communities centroids within 2 and 3 mile ring from the census block group centroid to form a new regression equation:

(9) $$\delta_h + \widehat{\alpha_{0\rho}}\,\rho_h = \alpha_{0X}X_h + \alpha_{0N}\widetilde{N}_h + \tilde{\xi}_h$$

where the tildes indicate the presence of additional control terms in the neighborhood variables vector. With these new variables, equation (9) is estimated using ordinary least square (OLS). By setting the OLS residual, $\tilde{\xi}_h$, equal to zero, the instrument for housing price is obtained as follows:

(10) $$\rho_h^{iv} = \frac{\delta_h - \widehat{\alpha_{0X}}X_h - \widehat{\alpha_{0N}}\widetilde{N}_h}{-\widehat{\alpha_{0\rho}}}$$

As mentioned above, the instrument price is dependent on the initial value of $\widehat{\alpha_{0\rho}}$. In order to eliminate this dependence, this study will apply the method used by Allen Klaiber and Phaneuf (2010). The strategy is that after determining the initial price instrument and running IV, the estimated price coefficient is obtained and the entire process of determining the price instrument is re-run using the new price coefficient as the initial guess. By repeating this process several times, the price coefficient eventually stabilizes and the initial dependence on the conjecture for the price coefficient is removed.

To provide some intuition for the relationship between parameters from the second stage estimation and traditional hedonic regression, I transform equation (8) as:

$$(11) \quad \rho_h + \frac{1}{\alpha_{0\rho}}\delta_h = \frac{\alpha_{0X}}{\alpha_{0\rho}}X_h + \frac{\alpha_{0N}}{\alpha_{0\rho}}N_j + \frac{1}{\alpha_{0\rho}}\xi_h$$

In equation (11), coefficient on δ_h, which is the inverse of the coefficient on price, provide the link between the first and second stage estimation. If all the households are homogeneous, equation (11) reduced to a hedonic price regression. Therefore, the mean indirect utility estimated in the first stage provide an adjustment to the hedonic price regression, and the second stage estimation of the sorting model returns more accurate mean preferences.

4. Data Sources

4.1 Housing and Household Characteristics

The study area in this study is the Franklin County of Ohio State. Analysis in this study requires data on house price and characteristics, environmental disamenities, and household characteristics. These data are not available at the same spatial scales, and data need to be merged from multiple sources. The housing data used in this study are real transaction data of residential one family dwellings of Franklin County, which can be obtained from the Franklin County Auditor's office. The data provide transaction records for residential properties located in Franklin County. Each record includes the property's address, transaction price and the structure characteristics. Using property address, each transaction can be located on the map of Franklin County, and neighborhood characteristics for each house can be obtained. Map location for each house is also used to calculate distance to the nearest landfill. Housing characteristics applied in this study include number of rooms, whether equipped with air conditioner and fire place, house age, distance to the nearest landfill and the type of the landfill.

Though housing transaction data provide detail information about the house, there is no information about households who are occupying the houses, and census data on individual household are not publicly available. As a result, aggregate information need to be used to approximate household characteristics. In this study, household characteristics and neighborhood characteristics are approximated using census data. Household characteristic variables include race of householder, household poverty status, household size, and education attainment. I collect 2010 census block data on the population of each racial group (black and white) and the average household size of each census block. Household income and education attainment are approximated using block group level data from 2006-2010 American Community Survey 5-year estimate. To analyze the effects of landfills on different household types, I stratify all the households based on race and poverty status. The hypotheses are that rich and white households are more likely to live further from the landfills than poor and black households. The mean value for percentage of white and black of all the census blocks is used to divide household race, and percentage of households with income under poverty line in the census block group is used to define household poverty status. For the sample, the average percentage of white is 73%, the average percentage of black is 18%, and the average poverty level is 16%. As a result, households live in census blocks with more than 73% of white are defined as white, while households live in census blocks with more than 18% of black. Households live in a census block group with poverty level higher than 16% are defined as poor. A householder is considered to have the high school degree if percentage of population with high school degree and higher in the block group that the household live is higher than the average value of the whole sample.

Therefore, household education attainment is described by a dummy which equals 1 if householder get high school degree and higher.

In this study, neighborhood is defined by census block group, and neighborhood characteristics include employment rate, ratio of black population, poverty level and school district quality. School quality is measured by school performance index, which is calculated based on student performance on the Ohio Achievement Assessments and Ohio Graduate Test. The summary of statistics of all the variables is shown in table 3.1. The summary statistics for housing characteristics indicates that the average house price is $155682, and the average number of bathrooms and bedrooms is 2 and 3 respectively. The summary statistics for households characteristics Table 3.1 also shows that the average household size of the data set is 2.7 and 53% of the households get high school degree or higher. According to household type classification, 67% of the households are classified as white, while 28% of the households are poor.

Table 3.1 Descriptive Statistics (Means)

Variable	Mean	Standard Deviation
Household characteristics		
White	0.66	0.47
Black	0.29	0.45
Poor	0.28	0.45
Education attainment	0.53	0.50
Household size	2.67	0.50
Number of observations	54462	

Housing characteristics		
House price	155682	98953
Bathrooms	1.90	0.76
Bedrooms	3.20	0.66
Fire place	0.48	0.58
Air conditioner	0.85	0.36
House age	43.08	100.50
Distance to the nearest landfill	8.03	3.80
Sanitary Landfill	0.05	0.22
Number of observations		54462
Neighborhood characteristics		
Population	1313	795
Percent of white	0.67	0.27
Percent of black	0.24	0.27
Poverty level	0.19	0.20
Employment rate	0.53	0.15
School quality	91.45	9.79
Number of observations		886

Note: All distances are measured in miles. Landfill type is a dummy variable which equals 1 if the nearest landfill is a municipal solid waste landfill, otherwise equals 0. Education attainment is an indicator whether the householder get high school degree or higher.

4.2 Environmental Disamenity

To investigate the impact of landfills on residential sorting, I consider distance to the nearest landfill and the type of the landfill as measures of environmental disamenities associated with landfills. Three landfills are included in this analysis. Franklin County Sanitary Landfill is a municipal solid waste landfill and accepts most household wastes. Franklin Road C&D Recycling Solutions and Scott Wrecking C&D Landfill are licensed strictly to accept demolition material. Distance to the nearest landfill is used to measure spatial environmental disamenity caused by landfill. In addition, different types of landfills may impact property values differently, and as a result I also analyze the effect of landfill type on households' residential location choice. The type of the nearest landfill is expressed by a dummy variable which equals 1 if the nearest landfill is a sanitary landfill which accepts municipal solid waste or household waste, while equals 0 if the nearest landfill is a construction and demolition landfill which accepts demolition material. Table 3.1 shows that average distance to the nearest landfill of all the households is 8.03 miles, and only 5% of the households live near a municipal solid waste landfill.

4.3 Choice Set

This study attempts to analyze household residential location choice in face of environmental disamenities by using a sorting model. Households choose their location from a discrete set housing alternative, and as a result determination of the housing choice set is very important for our analysis. In this study, I keep all the houses transacted during 2006 to 2010, and after cleaning the data, there are 54462 single family dwelling transactions in our data set. Following Tra (2013), this study assumes that each of the 54462 housing units chosen by the households in the sample represents a housing type. Though some papers (Tra 2010,Allen Klaiber and Phaneuf 2010) use discrete housing types rather than housing units to reduce the number of alternatives, Tra (2007) has shown that alternatives characterization of the product space using a smaller versus a larger number of housing types yields very similar parameter estimates. Therefore, the households' relevant choice set of alternatives are the 54462 housing types in the sample. However, the large choice set will make the estimation computationally infeasible. To solve this problem, following McFadden (1978) I construct the choice set by sampling a few alternatives from the full set of available alternatives, which includes the household's chosen residential location and a random sample of several nonchosen alternatives. This estimation strategy results in consistent estimates, but does reduce the precision of the first stage estimates.

5. Estimation Results

5.1 First Stage Estimation Results

The model is estimated at the level of house types, which are defined by housing units in this study. To characterize a choice alternative, all structural and neighborhood variables used in the model are created at the house type level. The purpose of the first stage is to recover the interaction parameters as well as a vector of mean indirect utilities for each housing type. In specifying the model, I include a limited set of interactions between household characteristics and housing and neighborhood attributes. By reducing the number of interactions according to the reasonable intuition, the degree of freedom for the estimation could be conserved and the potential problem of linearity could also be limited. Based on the criteria described previously, I stratify all the households into white, black and poor. The interactions estimated in the first stage include interactions of household size with the number of bathrooms and bedrooms. I also include interactions between household characteristics with neighborhood demographic composition to express household heterogeneous preference to live with households of same race. I also assumed that households with higher education would like to live in neighborhood with higher school quality, and this is tested by the coefficient on the interaction between household's education attainment and school district quality. To express the effect of landfill on household residential location choice in the first stage estimation, I interact distance to the nearest landfill with household race and household poverty status. In addition, different types of landfills may affect household's decision differently, and I also include interactions of household characteristics with distance to the nearest landfill and landfill type. The type of the nearest landfill is expressed by a dummy variable which equals 1 if the nearest landfill is a sanitary landfill, otherwise equals 0. The hypotheses here are that white and rich households are more likely to live in a house with longer distance to a landfill, and demolition landfills may impact property value more than a sanitary landfill.

The first stage estimation of the household-specific taste parameters are shown in table 3.2 shows. The results indicate that all the variables are statistically significant. Rather than the magnitude of the coefficients, I am more interested in the signs of the coefficients on the interactions. Coefficients on interactions between household and housing characteristics indicate that bigger households are more likely to live in houses with more bathrooms and more bedrooms. Coefficients on interactions between household race and neighborhood demographic composition confirm our assumption that households prefer to live with households of same race. Our interest is household heterogeneous preferences to environmental disamenity caused by landfills. The results indicate that poor households are more likely to select houses closer to landfills. Coefficients on interactions of household race with distance to the nearest landfill and landfill type show that keeping distance fixed, black households are more probably to choose houses located near a demolish landfill. Table 3.2 also shows that poor households prefer houses with lower price, and households with higher education attainment are more likely to sort into neighborhood with higher school quality. Generally, parameters in the first stage estimation indicate that there exists variation in the preferences for households for the characteristics of their housing choice.

Table 3.2 First Stage Estimation Results

Variable interactions	Parameter Estimate	Standard Error
Household size * Bathroom	0.074***	0.003
Household size * Bedroom	0.021***	0.003
Black * Percent of black	4.172***	0.037
White * Percent of white	2.528***	0.130
White * Percent of black	-2.757***	0.151
Poor * Landfill distance * Sanitary Landfill	-0.121***	0.013
Black * Landfill distance * Sanitary Landfill	-0.366***	0.037
White* Landfill distance * Sanitary Landfill	0.001*	0.005
Poor * House price	-0.1278***	0.002
Education attainment * School district quality	0.014***	0.001

Note: *** indicates statistical significance at 99 percent level; ** indicates statistical significance at 95 percent level; * indicates statistical significance at 90 percent level.

5.2 Second Stage Estimation Results

As mentioned before, when the choice set is large, it is computationally restrictive to estimate the fixed utility for each choice. To solve this problem, this study follows McFadden (1978) and constructs the choice set by sampling a few alternatives from the full set of available alternatives, which includes the household's chosen residential location and a random sample of several nonchosen alternatives. The method of contract mapping is used to estimate choice-specific constants that clear the housing market. Based on the estimation results of the household-specific taste parameters and choice-specific constants from the first stage estimation, the second stage estimation can be implemented. When estimating equation (4), one important underlying assumption is that housing prices are uncorrelated with unobserved characteristics of residential locations. However, there is likely significant correlation between housing prices and unobserved housing/neighborhood attributes. Therefore, I create an instrument for house price when estimating the regression. As mentioned previously, sorting model returns mean marginal willingness to pay more accurately than hedonic price regression, and to make the results comparable, mean marginal willingness to pay [13] for each housing and neighborhood characteristics from both sorting model along with the traditional hedonic price estimation are reported in table 3.3. Comparing the estimated mean preference for housing characteristics from sorting model and hedonic price regression, excepting the estimated mean preferences for air conditioner and house age, mean preference for other housing characteristics are overestimated in the hedonic price regression. The mean marginal willingness to pay for black neighbors from our sorting model is almost two times of that from the hedonic model which indicates notable difference. Our interests are the mean preferences relating to landfill. Table 3.3 shows that hedonic price regression underestimates the mean marginal willingness to pay for the distance to the nearest landfill, but when controlling for landfill type, the estimated mean preferences from both models are almost identical. For neighborhood employment rate, the mean preference from the hedonic price regression is not statistically significant. The mean preferences for school district quality and poverty level are overestimated in the traditional hedonic model.

[13] Implied mean willingness to pay for each neighborhood characteristics is estimated using the following equation: $\rho_h = -\frac{1}{\alpha_{0\rho}}\delta_h + \frac{\alpha_{0X}}{\alpha_{0\rho}}X_h + \frac{\alpha_{0N}}{\alpha_{0\rho}}N_j + \frac{1}{\alpha_{0\rho}}\xi_h$, where $\alpha_{0\rho}$ is the estimated coefficient on house price.

Table 3.3 Implied Mean Marginal Willingness to Pay (MWTP) Measures

	Implied mean MWTP measures (IV Estimation)		Hedonic price regression (OLS Estimation)	
	Parameter Estimate	Standard Error	Parameter Estimate	Standard Error
Implied coefficient on choice-specific constant	10.753***	0.002	-	-
Bathrooms	1.519***	0.029	5.699***	0.308
Bedrooms	0.188***	0.021	1.325***	0.255
Air conditioner	1.966***	0.037	1.071**	0.464
House age	-1.364***	0.015	-0.155	0.187
Fire place	1.728***	0.023	2.636***	0.276
Percent of black	-14.602***	0.066	-8.044***	0.790
Distance to the nearest landfill	0.468***	0.004	0.144***	0.049
Distance to the nearest landfill * Sanitary Landfill	-0.435***	0.010	-0.468***	0.127
School district quality	-0.093***	0.002	0.119***	0.021
Employment rate	1.013*	0.092	-0.532	1.154
Poverty level	-0.740*	0.098	-2.724**	1.228

Note: Implied mean marginal willingness to pay (MWTP) is calculated using estimated coefficient on price from the second stage IV estimation. House prices are measured in $10,000. *** indicates statistically significant at 99% and higher; ** indicates statistically significant at 95% and higher; *** indicates statistically significant at 90% and higher.

5.3 Heterogeneity in Marginal Willingness to Pay

Given the parameters estimated from both stage, we are able to interpret how household preferences for each housing and neighborhood attributes vary with household characteristics. Table 3.4 reports heterogeneous marginal willingness to pay measures for a subset of the variables in the second stage based on the interactions I created in the first stage[14]. Besides the mean household structure in our sample, marginal willingness to pay measures are also calculated for eight simulation scenarios for household structure based on four household characteristics. The results demonstrate the heterogeneity in preferences. For one more bedroom and bathroom the marginal willingness to pay increases from $20323 and $7043 for a household of two to $25484 and $12204 for a household of four respectively when the other household characteristics are the same. Black households would like to pay $2600 for one percent increase of the black population in their neighborhood, while white households would like to pay -$4425 for the same change. For marginal willingness to pay for distance to the nearest landfill, when keeping all the other household characteristics the same, rich and white households would like to pay more than poor and black households to move one mile further from the landfill. Table 3.4 also indicates that marginal willingness to pay for school quality increases from $935 for a householder without high school degree to $2473 for a householder got high school degree and higher. Existing literatures (Hite 2001, Kinnaman 2009) on marginal willingness to pay for landfill are general based on hedonic price regression, which did not capture how marginal willingness to pay differs among different household types. Our research fills this gap in the literature.

[14] The marginal willingness to pay values I calculate are conditional on the proxies used to measure household characteristics, and the average percent of households living near a sanitary landfill is used to obtain the marginal willingness to pay for distance to the nearest landfill.

Table 3.4 Heterogeneity in Marginal Willingness to Pay (MWTP) Measures

Panel A

Household structure	mean	Scenario 1	Scenario 2	Scenario 3	Scenario 4	Scenario 5	Scenario 6	Scenario 7	Scenario 8
Household size	2.67	2	2	2	2	4	4	4	4
Black	0.29	1	1	1	1	0	0	0	0
Poor	0.28	0	0	1	1	1	1	0	0
Education attainment	0.53	0	1	0	1	0	1	0	1

Panel B

Choice attributes changes	Heterogeneous MWTP ($)								
Bedroom (+1)	22052	20323	20323	20323	20323	25484	25484	25484	25484
Bathroom (+1)	8772	7043	7043	7043	7043	12204	12204	12204	12204
Percent of black (+1%)	-292	2600	2600	2600	2600	-4425	-4425	-4425	-4425
Distance to landfill (+1 mile)	3659	2441	2441	1790	1790	3763	3763	4414	4414
School district quality (performance index +1)	1750	935	2473	935	2473	935	2473	935	2473

Note: Coefficients used to calculate the heterogeneity in marginal willingness to pay (MWTP) are obtained from IV estimation. Panel A of the table shows the mean household structure of sample and 8 simulation scenarios of the household structure. Panel B of the table describe the heterogeneous MWTP associated with each of the simulation scenarios listed in the panel A.

5.4 Housing Consumption Measures

To explore how household heterogeneous preferences for landfills impact housing consumption, I did an equilibrium simulation using parameters estimated in the two stage equilibrium sorting model following Bayer et al. (2007). In the whole study, I use two schemes to stratify households, one for race and one for poverty status. Therefore, the focus of our simulation is how heterogeneous preferences respect to race and poverty status change affect household housing consumption, and during the simulation, I switch off heterogeneous disutility of landfills with respect to race and poverty status, and calculate a new set of prices that clears the market. Taking the new prices and the initial sociodemographic composition of each neighborhood, I calculate the probability that each household chooses each housing type and aggregate these choices to the neighborhood level to get the predicted demographic composition of the neighborhood. Then the initial neighborhood demographic composition is replaced by the predicted demographic composition, and new equilibrium is calculated. This process is continued until neighborhood demographic characteristics converge. Using the new equilibrium probability, the weighted-average consumption measures for households in a given race or poverty status stratification are constructed. Difference between sample housing consumption patter and weighted-average consumption pattern describe the role of household heterogeneous preference in shaping the distribution of housing consumption.

The sample housing consumption patterns[15] and simulated consumption patterns for household race and poverty status measures are shown in table 3.5. Panel A describes sample average consumption levels of each housing characteristics for households of different race and poverty status, while panel B gives average consumption levels of each housing characteristics

[15] The sample consumption measures for housing are computed simply by averaging housing consumption over all households in a given stratification.

after switching off heterogeneous preferences for landfills. When turning off landfill preferences, there is little impact on housing consumption of white. For black and poor households, I see more notable changes on the housing consumption. After deleting the importance of landfill on household residential location choice, the black and poor are more likely to select smaller houses. This confirms that landfill did discount nearby house value. If a landfill is closed, the price of existing houses will increase and black and poor households will not be able to afford houses as large as they are living now. Also, in the counterfactual simulation, black and poor households are more willing to buy newer houses.

Table 3.5 Housing Consumption Measures

Panel A: Existing consumption patterns					
	Bathroom	Bedroom	Air conditioner	House age	Fire place
Black	1.638	3.100	0.740	47.864	0.285
White	1.992	3.265	0.903	38.667	0.565
Poor	1.452	2.980	0.684	52.747	0.319
Panel B: Switch off landfill preferences					
	Bathroom	Bedroom	Air conditioner	House age	Fire place
Black	1.209	2.315	0.549	34.239	0.267
White	2.020	3.347	0.914	40.246	0.608
Poor	1.002	2.067	0.484	35.391	0.222

Note: This table gives average consumption levels of the column characteristics for households in the first row category.

6. Conclusions

Sorting and heterogeneity of households has been of subject of many empirical studies. In my study, I use a two stage equilibrium sorting model to analyze household sorting and environmental disamenities in the neighborhood of Franklin County of Ohio State. Environmental disamenity is an important factor affecting household residential location choice, and our interests are the effects of disamenities associated with landfills on household sorting. Three landfills in the Frankly County of Ohio State are included in this study, and environmental disamenities are measured by distance to the nearest landfill and landfill type. To explore how household heterogeneous preferences for landfills affect housing consumption, I did an equilibrium simulation based on the full set of estimates of the equilibrium sorting model following Bayer et al. (2007). In the simulation I switch off heterogeneous disutility of landfills with respect to race and poverty status, and this simulation provide direct estimation of preference changes influencing households sorting. Because of the limitation of housing transaction data on household information, census block and census block group data are used as proxies to household characteristics.

Our empirical analysis indicates that households are heterogeneous in their willingness to pay for housing and neighborhood characteristics, which shaped the way that households sort across neighborhood. Based on results from the first stage estimation, poor households are more likely to select houses closer to landfills, and black households are less probably to choose houses located near a sanitary landfill when keeping distance fixed. To account for the potential price endogeneity problem in the second stage, an instrument variable for price is created based on the exogenous attributes of distant neighborhood. Since sorting model captures impact of household heterogeneity, it returns mean preference for housing and neighborhood characteristics more accurately than the hedonic price regression. By comparing mean marginal willingness to pay for distance to landfills from the two models, I conclude that hedonic price regression underestimated the result. When turning off landfill preferences, there is little impact on housing consumption of white, but for black and poor households, I see more notable changes. After deleting the importance of landfill on household residential location choice, black and poor households are not able to afford houses as large as they are living now.

References

Allen Klaiber, H., and D.J. Phaneuf. 2010. “Valuing Open Space in A Residential Sorting Model of the Twin Cities.” *Journal of Environmental Economics and Management* 60(2):57–77. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0095069610000574 [Accessed December 19, 2014].

Anselin L. and Gracia L. Nancy. 2008. “Errors in variables and spatial effects in hedonic house price models of ambient air quality”. *Empirical Economics*. 34:5–34.

Banzhaf, H.S., and R.P. Walsh. 2008. “Do People Vote with Their Feet ? An Empirical Test of Tiebout ’ s Mechanism.” *The American Economic Review* 98(3):843–863.

Bayer P., Robert M. and Kim R. 2003. “An Equilibrium Model of Sorting in an Urban Housing Market**:** The Causes and Consequences of Residential Segregation”. *Economic Growth Center,* Yale University: Working Paper.

Bayer, Patrick Robert, McMillan Kim, R. 2004. “An Equilibrium Model of Sorting in An Urban Housing Market.” *NBER woking paper seties* No.10865.

Bayer P., Ferreira, F.V., and McMillan, R. 2005. “Tiebout Sorting, Social Multipliers and the Demand for School Quality”. *NBER woking paper seties* No. 10871.

Bayer, P., F. Ferreira, and R. McMillan. 2007. “A Unified Framework for Measuring Preferences for Schools and Neighborhoods.” *Journal of Political Economy* 115(4):588–638.

Bayer, P., F. Ferreira, and R. Mcmillan. 2007. “A Unified Framework for Measuring Preferences for Schools and Neighborhoods.” *Journal of Political Economy* 115(4):588–638.

Bayer, P., N. Keohane, and C. Timmins. 2009. "Migration and Hedonic Valuation: The Case of Air Quality." *Journal of Environmental Economics and Management* 58(1):1–14. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0095069609000035 [Accessed December 17, 2014].

Bayer, P., and R. McMillan. 2012. "Tiebout sorting and neighborhood stratification." *Journal of Public Economics* 96:1129–1143.

Berry, S.T. 1994. "Estimating Discrete Choice Models of Product Differentiation." *The RAND Journal of Economics* 25(2):242–262. Available at: http://www.jstor.org/stable/2555829.

Biying Y., Junyi Z. and Akimasa F.2012. "Analysis of the Residential Location Choice and Household Energy Consumption Behavior by Incorporating Multiple Self-selection Effects." *Energy Policy* 46: 319-334.

Bouvier, R.A., J.M. Halstead, K.S. Conway, and A.B. Manalo. 2000. "The Effect of Landfills on Rural Residential Property Values : Some Empirical Evidence." *The Journal of Reginal Analysis and Policy* 30(2).

Brueckner, J. K., Thisse, J. F. and Zenou, Y. 1999. "Why is Central Paris rich and downtown Detroit poor? An amenity-based theory". *European Economic Review*, 43: 91–107.

Bui, L.T.M., and C.J. Mayer. 2003. "Regulation and Capitalization of Environmental Amenities: Evidence from the Toxic Release Inventory in Massachusetts." *The Review of Economics and Statistics* 85(3):693–708. Available at: http://www.mitpressjournals.org/doi/pdf/10.1162/003465303322369821\npapers3://publication/uuid/6D14268D-9A44-4FAE-8399-B0E863027326.

Case, K. E. and Mayer, C. J. 1995. "Housing price dynamics within a metropolitan area". Federal Reserve Bank of Boston. Working Paper No. 95-3.

Chay K.Y. and Greenstone M. 2005. “Does air quality matter? Evidence form the housing market”. *Journal of Political Econonics*. 113(2):376-424.

Claudio Frischtak and Benjamin R. Mandel. 2012. “Crime, House prices, and Inequality: The Effect of UPPs in Rio.” Federal Reserve Bank of New York Staff Reports: No. 542

Cullen J.B. and Levitt S.D. 1999. “Crime, Urban Flight and the Consequences for Cities.” *Review of Economics and Statistics,* 81:159-69.

Crowder, K., and L. Downey. 2010. “Interneighborhood Migration, Race, and Environmental Hazards: Modeling Microlevel Processes of Environmental Inequality.” *American Journal of Sociology* 115(4):1110–1149.

Decker, C.S., D. a. Nielsen, and R.P. Sindt. 2005. “Residential Property Values and Community Right-to-Know Laws: Has the Toxics Release Inventory Had an Impact?” *Growth and Change* 36(1):113–133. Available at: http://doi.wiley.com/10.1111/j.1468-2257.2005.00269.x.

Duijn M. and Rouwendal J.2013. “Cultural Heritage and the Location Choice of Dutch Households in a Residential Sorting Model.” *Journal of Economic Geography*, 13:473–500.

Gibbons S. and Machin S. 2008. “Valuing School Quality, Better Transport, and Lower Crime: Evidence from House prices.” *Oxford Review of Economic Policy,* 24(1): 99-119.

Hite, D. 2000. “Environmental Equity.” *Growth and Change* 31(1):40–58.

Hite, D. 1998. “Information and Bargaining in Markets for Environmental Quality.” *Land Economics* 74(3):306–316. Available at: http://doi.org/10.2307/3147114.

Hite, D. 2001. “Property-Value Impacts of an Environmental Disamenity : The Case of Landfills.” *Journal of Real Estate Finance and Economics* 22(2/3):185–202.

Ihlanfeldt K. and Mayock T. 2010. “Panel Data Estimates of the Effects of Different Types of Crime on Housing prices.” *Regional Science and Urban Economics,* 40:161–172.

Kinnaman, T.C. 2009. “a Landfill Closure and Housing Values.” *Contemporary Economic Policy* 27(3):380–389. Available at: http://doi.wiley.com/10.1111/j.1465-7287.2008.00143.x [Accessed December 27, 2014].

Kuminoff, N. V, V.K. Smith, and C. Timmins. 2013. “The New Economics of Equilibrium Sorting and its Transformational Role for Policy Evaluation.” *Journal of Economic Literature* 51(4):1007–1062. Available at: http://www.nber.org/papers/w16349.

Mastromonaco, R. 2015. “Do Environmental Right-to-Know Laws Affect Markets? Capitalization of Information in the Toxic Release Inventory.” *Journal of Environmental Economics and Management* 71:54–70. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0095069615000194.

Lynch A. and Rasmussen D. W. 2001. “Measuring the Impact of Crime on House prices.” *Applied Economics,* 2001:1981-1989

McFadden, D. 1973. *Conditional Logit Analysis of Qualitative Choice Behavior* P. (Ed. . Zarembka, ed. New York: Academic Press.

McFadden, D. 1978. *Modeling the choice of residential location.pdf* In Spatial. and J. W. (eds. . A. Karlqvist, L. Lundqvist, F. Snickars, ed. North Holland: Amsterdam.

McFadden, D. 1978. *Modelling the Choice of Residential Location* Krlqvist, A. et al. (Ed.), ed. New York: Elsevier North Holland. Available at: http://cowles.econ.yale.edu/P/cd/d04b/d0477.pdf.

Ready, R.C. 2010. “Do Landfills Always Depress Nearby Property Values?” *Journal of Real Estate Research* 32(3):321–339.

Rosen, S. 1974. “Hedonic prices and implicit markets. Product differentiation in pure competition”. *Journal of Political Economy,* 82: 34-55.

Sampson J.D. and Wooldredge. 1986. “Evidence that High Crime Rates Encourage Migration Away from Central Cities.” *Sociology and Social Research,* 90:310-14

Sieg, H., V.K. Smith, H.S. Banzhaf, and R. Walsh. 2004. “Estimating the General Equilibrium Benefits of Large Changes in Spatially Delineated Public Goods.” *International Economic Review* 45(4):1047–1077.

Tiebout, C.M. 1956. “A pure theory of local expenditures.” *The Journal of Political Economy* 64(5):416–424. Available at: http://www.jstor.org/stable/1826343.

Tra, C.I. 2010. “A discrete choice equilibrium approach to valuing large environmental changes.” *Journal of Public Economics* 94(1-2):183–196. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0047272709001212.

Tra, C.I. 2007. *Equilibrium Welfare Impacts of the 1990 Clean Air Act Amendments in the Los Angeles Area*. College Park: University of Maryland.

Tra, C.I. 2013. “Measuring the General Equilibrium Benefits of Air Quality Regulation in Small Urban Areas.” *Land Economics* 89(2):291–307.